

JOURNAL Submitted (06-DEC-2000) Plant and Microbial Biol. Dept., University of California, Berkeley, 361 Koshland Hall, Berkeley, CA 94720, USA

FEATURES

Location/Qualifiers
1. 3896
/organism="Oryza sativa"
/db_xref="taxon:4530"
241-3414
/note="EMF1"
/codon_start=1
/product="embryonic flower 1-like protein"
/protein_id="AK98529.1"
/db_xref="GI:15430699"

CDS

BASE COUNT 1175 a 791 c 935 g 995 t
ORIGIN

Query Match 99.9%; Score 3892.8; DB 8; Length 3896;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3894; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 CGCGGCTGTCGAGCAACCAACCCCAAGTGTCTAGTGTGTCAGCGAGTACT 60
1 CGCGGCTGTCGAGCAACCAACCCCAAGTGTCTAGTGTGTCAGCGAGTACT 60
61 GATTGATGCTTCTGATATATCCAGAGCTGCTGTTGTTGTTGTTGTTGTT 120
61 GATTGATGCTTCTGATATATCCAGAGCTGCTGTTGTTGTTGTTGTTGTT 120
121 TGTCCTTGAGATTGTGATGCTAATTCGCGGCTTACAAAGTACAGTGTGATAT 180
121 TGTCCTTGAGATTGTGATGCTAATTCGCGGCTTACAAAGTACAGTGTGATAT 180
181 TGAATTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
181 TGAATTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
181 TGAGTGTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
241 ATGAGATTGTTGCAAGTAGATCAGAGGAGAGCTGTTGTTGAGAGAACTGATCTT 300
241 ATGAGATTGTTGCAAGTAGATCAGAGGAGAGCTGTTGTTGAGAGAACTGATCTT 300
301 GCTGCTGAGGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
301 GCTGCTGAGGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
301 GCTGCTGAGGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
301 GCTGCTGAGGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
361 GATGAGCGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
361 GATGAGCGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
421 ATAAAGGATATGCTTCTTCTCAGAAAGATCCAAATTCCTGCTCTATCTCGG 480
421 ATAAAGGATATGCTTCTTCTCAGAAAGATCCAAATTCCTGCTCTATCTCGG 480
481 ATTTCATGACCAAGAAATGTCAGCAACAAGCTAGTTCAGCCATTTCTGTA 540
481 ATTTCATGACCAAGAAATGTCAGCAACAAGCTAGTTCAGCCATTTCTGTA 540

481 ATTTCATGACCAAGAAATGTCAGCAACAAGCTAGTTCAGCCATTTCTGTA 540
541 GCAAGTTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
541 GCAAGTTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
541 GCAAGTTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
601 GGAAGAGCAACCAAGACTCTCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 660
601 GGAAGAGCAACCAAGACTCTCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 660
601 GGAAGAGCAACCAAGACTCTCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 660
601 GGAAGAGCAACCAAGACTCTCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 660
661 AATTGTTGCGAGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
661 AATTGTTGCGAGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
721 ACAAATCACTCAAG 780
721 ACAAATCACTCAAG 780
721 ACAAATCACTCAAG 780
721 ACAAATCACTCAAG 780
781 GCAATGACTCCAAATGCAATGCGCTTCTGCAAGAGAGAGAGAGAGAGAGAG 840
781 GCAATGACTCCAAATGCAATGCGCTTCTGCAAGAGAGAGAGAGAGAGAGAG 840
841 GATTCACCAATGAAAGATTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
841 GATTCACCAATGAAAGATTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
901 TCTGAGAGCAACACTCTGTTGATGTTGAGGCTTACCTGGAAGTTCCAGATGG 960
901 TCTGAGAGCAACACTCTGTTGATGTTGAGGCTTACCTGGAAGTTCCAGATGG 960
961 CACATGAGCAACACTCTGTTGATGTTGAGGCTTACCTGGAAGTTCCAGATGG 1020
961 CACATGAGCAACACTCTGTTGATGTTGAGGCTTACCTGGAAGTTCCAGATGG 1020
1021 CTTCAAAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1080
1021 CTTCAAAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1080
1081 AATTGACCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1140
1081 AATTGACCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1140
1141 TGAATTTGACCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1200
1141 TGAATTTGACCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1200
1201 GAGCATTGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1260
1201 GAGCATTGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1260
1261 AAGTTGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1320
1261 AAGTTGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1320
1321 GGAAGAGCTTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
1321 GGAAGAGCTTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
1381 AACCAGTTGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1440
1381 AACCAGTTGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1440
1441 GAGGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1500
1441 GAGGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1500
1441 GAGGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1500
1501 AACCATACAGTGGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1560
1501 AACCATACAGTGGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1560
1561 TCTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1561 TCTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620

OY	1621	acgygaagctgltgcatcaacagctgtgtctalccagctggygaatttggagcaacaaanaagtg	1660
Db	1621	ACTGGAAGTGTGCTACACAGATGGCTCATCCAGCTGGGAATTTGAGCAACAAAAAATGTG	1680
OY	1661	acaccacactcgagttactgacgtatgtatgtgaatgatgacccaanaatggtcttgacaa	1740
Db	1661	ACACCACCTGGAGTACTGACATGTATGTATGTAAGTATCTGAAAATGTCCTTGACACA	1740
OY	1741	aatatgtacaaagacagaatgtctgtlccagcatgtatcagaanaatctccacaagaagtgtcca	1800
Db	1741	AATATGCATTAAGACAGATGTCTGTGCAGCATGTATCGAANAATCTCCACACAGAGGTGTCA	1800
OY	1801	tcaaaaggggaaaaaacagcgggttltgagtaaggggaaacaacatctagctgtcgtatgccaa	1860
Db	1801	TCAAAGGGGAAAACAGCGGGTTTGAGTAAGGGGAAAACATTCAGCTGTCTATGACAAA	1860
OY	1861	latgtgtgtgaagcaccacgaanaatgtgtcagaacatacatgtatctagcgcgaagaatcaa	1920
Db	1861	TATGTGTGTGAAGCACACGAATAATGGTTCAGAACATTCATGTACTCAGCGCAAGAAATCAA	1920
OY	1921	ttccagatgtgaaacccgaaacactctgttctgagtlactcgtggcaaaagtttctccagctgag	1980
Db	1921	TGCCAGATGTGAACCCGAAAACCTGTTCTGAGTCATCTCGGCAAAAGTTTCTCCAGCTAG	1980
OY	1981	catgatctccaaatgatgtcttgacaccttcataagcagaagctctcgaagaagaanaagag	2040
Db	1981	CATGATATCCAAATTTATGTGTGACTTTCATGTAGCAGAGTCTACCCAGAGAAAAAAGAG	2040
OY	2041	caaaaacttgaaagtgactcgttgaaaaaacagaccatgatagatgatccccaatgtgataat	2100
Db	2041	CAAAAACTTGAAGTACCTCTGTAAAAACAGACCATGATGTATGATGATCCCTCGATATTT	2100
OY	2101	gttgaactctgactgaagtaaaaaacagcactgtgaaggaagacctaatactcgaagactgattgtct	2160
Db	2101	GTTGAACTGCTAGCTTAATAAACCGCATGAGAGCACTTATACCTGAGCTGATTTGTCT	2160
OY	2161	gacatcaaaccgatltcaatccaaagaacactgtcatgatgatgtgtataagtaagtctgcc	2220
Db	2161	GACATCAACCGTATTTCAATCCAAAGACAACGCTGATGATGATTTGTGTAATAGTACTGCC	2220
OY	2221	aaggaatggttcaagttatgtgataaagttgtgttgaacactaattcccaagaagaagctctg	2280
Db	2221	AAGGATGGTTCAGTTATGTGATCAAGTATGTTGTGACATCAATTTCCCAACGAAAGTCTTG	2280
OY	2281	gcatcccaagaatlaacacagaagaaggttacaagggctcatlttgcatltpacacaaagaagctc	2340
Db	2281	GCATCCCAAGTACACAGAAAGGATTACAGGGCTCATTTGGCATTTGACACACAAAGTCT	2340
OY	2341	cccaatctctcagaactctcaagcttaetcaaggaacagcagaacaaattctcgatgtgaaga	2400
Db	2341	CCCAATCTCTCGAACTTTGAGTCTACTCAGGAACACACAAACAATTTGGGTGTGGAAGAA	2400
OY	2401	atgtgtcaactatgtgtgaagcttcaacaaatlttcaatcatgatgatgaatgaatattgtct	2460
Db	2401	ATGTGTCACTATTTGCTGCAACCTCACCACTATTTTTCACATCATGATGATCAGTATTTCT	2460
OY	2461	gaagacccaactagacatltggyggccgttaaagacgcgaagaagactaagctltggygaacatt	2520
Db	2461	GAAACACCACTGTGAACATTTGGGGCCGTAAAGAGCGCAAAAGACTTAACGTHGGGCAATTT	2520
OY	2521	aagggcaactcaagaanaatttctccaagcagaacaatgtgtgtgtccaatlttagaactgtgtatc	2580
Db	2521	AAGGCCACTTACAAGAAATTTCTCCAGACGCAACATGTGTGTCTCATATTTTAAGACTGTGTATC	2580
OY	2581	caagcagtttgactgtgaacttctacatgtcatatgtgaggtcttccagaaatltatgtcatctgc	2640
Db	2581	CAAGAGTTGACTTGTGACTTCTACTCATGTCTATGATGATTTCCAGCAATTTATGCACTCTGC	2640
OY	2641	caaccagtaattctgcacacttgagacgctatgtctgaagaagcggcttaaccaggttccatga	2700
Db	2641	CAACCAAGTAATTTGGCCACTGTGAGACCGTATATGTCTGAAGAAGCGGTTAAACGAGTCTCATCA	2700

QY	2701	agaaattcttccaagcacataagcaaccaatggaagcagtaaaatctaatctatctgtatcgcgagaat	2760
Db	2701	AGAAATTTTCCAAACACAAATAAGCAACCATGGAAGCGAGTAAGTATATGATCGGAGAAAT	2760
QY	2761	gctggaagaagtagctctgtatccctaaagaaatcaatgcctggagacatctctgagaatg	2820
Db	2761	GCTGGACAAAGTAGTCTGTATATCCTAAGAAATCCATCCCTGCACACGATCTTCTGGAATG	2820
QY	2821	atggtaccataaataatagaaagctctcccacaatacgtgaactctcagaaagacaatg	2880
Db	2821	ATGGATTCATACATATAGCAAGCTTCCCAACTATAGAACTTTAGACAGAAACAGATG	2880
QY	2881	gagttccaactcaaatctcgcgatgacataatcagtaacaaagatcaacacagaca	2940
Db	2881	GAGTCTCAACTTCATTAATTCATCGATATGACATATAACATAAAGATCAACACGACA	2940
QY	2941	tcaatctgcagtaaccctggaatgtgaagaatcccatctgaactctgaaagactttcaagat	3000
Db	2941	TCATATGGCACTAATCACTGGAATGGAATAATCCATTCATCTGAAGACTTATACAGGAT	3000
QY	3001	cagctgcatactgctgcacaacaccttcgcgccacatccagagatctgtgtgtctgtcc	3060
Db	3001	CAGCTGCATATCTGCAACAGACCTTTAGGCCCACTCCATAGATTGTGTGTGCTTGCTCC	3060
QY	3061	ctgtctgcagaagaatctgcaactgtgtcggaagactgtgcacaaatctgtgtataag	3120
Db	3061	TTGCTGCAGAGGAATAATGCAACTGGTGCGAATCTGGCACAAATCTGTATATAG	3120
QY	3121	ttagaagtgtaacaagaataaactcgcatacagataaacaagaagaacatcttgaac	3180
Db	3121	TTAGAGAGTGTAACAAGAAATACATCCCATCAATACACAGAAAGAACATTTGTAAC	3180
QY	3181	ctgaattctcgaatcttctcagcaaaatgtaatctgacatgtgtgtctgtctgaacc	3240
Db	3181	CTGAATTTCTGGAATGTTTTCAGCAAAATGGAATGCAATTCAGTTGGTCTGTATGCTCC	3240
QY	3241	agtcgacatcttctatcaagcagaagacagcatagctcaatcttggaacaggaagaagt	3300
Db	3241	AGTCGATTTTTTTATCATCAGCAGAGAACAGCATAGCTCAATCTTGGCCAGAGCAAGGT	3300
QY	3301	aaaaatggtcaatcccttggaatcggttgtgagaagaatatactgataaactaaacaagac	3360
Db	3301	AAAAATGTTCAATCCCTTGGATCGGTTTGTAGACAGAGATATCTGATATATCAACAAGAC	3360
QY	3361	ccaagtctatcttactacaatcagtaacgaataaagaatatactgataccgctggaagcaga	3420
Db	3361	CCAGCTGATTTTACTACATAGTAAGTAAGATATATGATTAATCCGCTGAAGCAGA	3420
QY	3421	aagtgtgtgtcataatctctgtaacatttacaatcataatcttataatgtggccaaa	3480
Db	3421	AAGTGTGTGTGATTAATTCCTGAACATTTACATATCAATTCATCTTATATGGCCCAAA	3480
QY	3481	tagtcaatctgtaaaggaagaggtcttgctgtgatactgtctgtaaggtctcttgaaatgtg	3540
Db	3481	TAGTCAATCTGTAAAGAGAGGCTTCTGGAATCTCTGTGAAGCTTCTTGTGAATTTGTG	3540
QY	3541	gattgcccaatcttcggaatggaacctgcgcagacagtgaaacaagggctctgcaagtgc	3600
Db	3541	GATGCCCAATTTTCTGGAATGGGAACCTGCCAGACATGATCAANAAGGGCTTTGCAAGGTCA	3600
QY	3601	gataccggattcttgcttctgcacagtcacaagaacgctccctcttactctgtatgttact	3660
Db	3601	GCATCGGTTTTTGTGTTTGGCAGTCCAAAGAAAGTCTCTCTGTACTTGTAGTTGTACT	3660
QY	3661	catactagctgcgtctgttctgacaaggagaatgttlaacctgtgtgaaaaaatgtctc	3720
Db	3661	CATACTAGCTGCGTGTGTTGTGACAAAGAGAAAGTGTAACTTTGTGAAAAAAAGTCTC	3720
QY	3721	ccccatttgaatatacaataaagaagttataatgttttgaagctgtgtgtgacgaag	3780
Db	3721	CCCCATTTTGTAAATTTACCATTAAGAGAGGTTATATAGTGTGTGAGCTGTGTGTACAGAGG	3780
QY	3781	cgaaagaatgtttctgtcgtgtttaaagtctbaaacgcataagctctcgttcaaatgtctg	3840

Db 3781 CGAGAAATGTTTGTGCTTAAAGTTGAAGACTAGCTCCGTATCAATGTCCTG 3840
 QY 3841 taactctagatgattgattgattacacttctgaagtcacacccggagaatttac 3896
 Db 3841 TAACTCTAGATGATTGATTGATTACCTTACTCTTGAAGTCAACACCGAGAAATTTC 3896
 RESULT 2
 AP001859 150594 bp DNA linear PLN 27-MAY-2000
 LOCUS Oryza sativa genomic DNA, chromosome 1, PAC clone: P0485D09.
 DEFINITION AP001859
 ACCESSION AP001859.1 GI:7630233
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa (cultivar: Nipponbare) DNA, clone: P0485D09.
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 150594)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone: P0485D09
 Published Only in Database (2000) In press
 2 (bases 1 to 150594)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (19-Apr-2000) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@da.affrc.go.jp,
 URL: http://www.dna.affrc.go.jp:82/, Tel: 81-298-38-7441,
 Fax: 81-298-38-7468)
 COMMENT
 The orientation of the sequence is from T7 to SP6 of the PAC clone.
 Genes were predicted from the integrated results of the
 following: GENSCAN 1.0, BLASTN 2.0, BLASTX 2.0 as well as
 SplicePredictor (October 1998 version). The genomic sequence was
 searched against the non-redundant database NRP (PIR, SWISSPROT,
 GENPEPT, PDB) from MAF DNA bank and the cDNA sequence database at
 RGP. Protein similarities of the coding regions were searched
 against NRP with BLASTP 2.0. ESTs represent the identified cDNA
 sequences using BLASTN 2.0 with the corresponding DDBJ accession no.
 and RGP clone ID.
 This sequence of this clone has an overlap with P0431F01 clone,
 DDBJ: AP001550 at the 3' end. This clone ends at the position 57,376
 of P0431F01. Detailed information on overlap and assembly quality
 together with annotation of this entry at
 http://www.dna.affrc.go.jp:82/genomicdata/Genomefinished.html.
 FEATURES
 Source
 1. 150594
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0485D09"
 join(7919..7964,8092..8280,8453..8473,8558..8610)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BA94767.1"
 /db_xref="GI:7630233"
 /translation="METRIFTIOVRAPELSVDNRIALFLESVGEIRDSAFVSPDSIRK
 HRRVLAFLIPEVDIRRRVLAIFYGVVYASMSYLVGCFRIGHVSPNPGRIRLOQIKAK
 "
 CDS
 complement(9235..9582)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BA94768.1"
 /db_xref="GI:7630235"
 /translation="MTARFAVEGERREAPAGRSMEERGCGGARRRRCASPAATAATRIGG
 ARRRGREGKRGVVDYLAIFPGWFLQKRGREGARGCEMDGEMARVGSARSGPSV
 STSAFVSREGNP"
 join(10733..10850,10933..10987,11596..11653,12285..12379),

12471..12574,12958..13086,13493..15302)
 /note="Similar to Arabidopsis thaliana chromosome 4, BAC
 clone Fd11; putative myb protein. (AL022537)"
 /codon_start=1
 /protein_id="BA94769.1"
 /db_xref="GI:7630236"
 /translation="MTSNGKAPKDGSPGSPAPQEGEISNEPKRRRLNGRTTGP
 TRSRKGMWTPBEDAIIISRAVOTYNGKMKWKIACFPDRDVOCLHMRQVLPDELK
 GWSKDEEITVQYVKNKIGPKKSTIAQALDGRIGKQCRERWVHNLNPGINKEMTDE
 BEITLIHARMYGNKMAELTFLEKRDINSLTKNNMNSVKKKRVNSYSGLLNQVSL
 PLNESANCNSSPAMTQONSBDSCGFARVENSNGSOSLAKVSCSQVHDTTPVPLG
 CDIOYANFMDKNEADHDSQSMGPQACVSAEVAASALPVAHVCHVSSNLDDQLQED
 PBOGILNMTDEMTPTSPFADNOCVCEIEHENSLEPLVDVAMEVLSMLSDSGAQD
 KIHFESEADNPNSNCSVWGLADDSQSPOMVSPCLISNMATDAPEDNRPGRKEMPLSQ
 FLAPYMPQSTNSSSVWGLADDSQSPOMVSPCLISNMATDAPEDNRPGRKEMPLSQ
 AAVVYSSSSSDAEFANPSCSNDHRVPSSTMESIPCCGOOYTNNAEERPALEKPE
 SLTQSVTAPDEQDKGALFEPPEPESLDVPPVSCDLVTSGLQDFSPGLGTPQLHSTMT
 NYCTPMRLMSPETTHDESTGVLLKSAKSFICTPSTLKKRHDLSLPDKRTIEKRYGT
 EKDRCVSDTSSTGIQSCINATKDALITTVLRIERSASRSKLRKLVFSDENKENSIG
 YTTPECKQSGSAGNDEHDEQTERSATVATVDLSGNLVSTSSP"
 complement(join(18608..18823,19905..20031,20072..20220,
 20774..21113,21407..21738,21975..22508))
 /note="ESTs AU082761 (S5084), D42006 (S5084) correspond to a
 region of the predicted gene.
 Similar to Caenorhabditis elegans chromosome 3, cosmid
 C17G10; alcohol dehydrogenase/riboitol dehydrogenase.
 (U28739)"
 /codon_start=1
 /protein_id="BA94770.1"
 /db_xref="GI:7630237"
 /translation="MAAABEGGGGGGAVDPSPVYTFVAMHTRHYRDKVYE
 IHOMYEGFORLSDRLFRDAPWPAEVAVSPDGDHVFLLTRELMTAHYARASSS
 SSAPLTPAQARASMANVCDLFSVYLGVVNNQMLNOMLIDNDEVFVQFSQCYRA
 KLRKSGDDELHOLKOPDKAMNYVILQALVERKSMALQIEREKELDEFTADYD
 DYCGSNVLMKGYMIGLIRHCLLDGFRFGCLAPIDNQQGVYTVIGSHISA
 IYHGRANLMPRYAARIREPKLILYIKYQYHOSPOYQIILKKEOMAYFLAVC
 LSLCPHNLIDENVSTQLEKRYKNDKMTNRDEEYAYVDLFTYACKKFTPEPPA
 LDLPITNYNDAYRLQKLFYHVKQOQLSGIRSKYKSTLYKLYIAQVNEVEAT
 LRAMKNI.SQYLYRSILMTYKHKHAAVDNNGKIVSSADPFYIKETCSSIRPKLSKN
 LAIHIOGVKTYRKLSSLSFSYFFFGFIEIGSGVSEIEYRNKMSNEPAPDIR
 join(23045..23145,24592..24670,25118..25216,25302..25415,
 25608..25750,26364..26445,27073..27158,27527..27760,
 29332..29917)
 /note="Similar to Arabidopsis thaliana chromosome 2 BAC
 T19C21; unknown protein. (AC004683)"
 /codon_start=1
 /protein_id="BA94771.1"
 /db_xref="GI:7630238"
 /translation="MAGSMVDRATSDMLIGPDWAKNMEICDICNRDPGOSKDVYKAL
 KRICHKRPKQIITATLLETAIRKNGDIFEMHYAEEDVLIHEMKIYKRSKSDNVKEXY
 LTMIDTWQENFEGFRARYROYVAAHYVAGAPAFRSRSPALPLFNGOSPAGRNRS
 PDQODEASSAGNDFPALMSSEIONANGINDVLAEMNLADPNREGLEOEYVELVD
 QCRTYKQRYVLLVNAATDEELMSQGLANDLQVRLAKHDALIAAGIAYVEKKRKSLQ
 ALVDEDSMDQSKKEQALVVIDPTQIENKRPSSASVOSFEDUALDAPFVNSGS
 ATPAPKSDLVGLDLSMDSTSTENSIALVHPVDPVDSVPSQNALAIADVTSQNTT
 NTKAKPADPFGVNSSSTLPGSPYNAPTRQLQAQDQPOQGVLPNGAGVNPSTSYPT
 SSGWNGQIANNATPPAQGTNYGAIYLYISOPFYSYFTIIPILGQNNACPLCSV"
 30166..30675
 /note="ESTs AU031545 (E61825), AU031546 (E61825) correspond
 to a region of the predicted gene.
 hypothetical protein"
 /codon_start=1
 /protein_id="BA94772.1"
 /db_xref="GI:7630239"
 /translation="MYNNQPGAMQPSOPAVTOMQSGFANQFGSMPPSHMQPQM
 QPSPGQAQPVMTAQPPMMGQFAMQPPRMQSOYRLVQOQAAQYASNSQR
 PTVYAGNMLDSQKMYGLSMDSSYTMGNMSSYSTTPSSSSSMQPSKPEDKLIGDLSL
 IAKTKONKA"
 complement(join(34629..34795,36831..36867,38803..38895,
 41780..41830))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BA94773.1"

CDS

CDS

CDS

CDS

CDS

Query Match Best Local Similarity 87.4%: Score 2952.2; DB 8; Length 150594; Matches 3487; Conservative 0; Mismatches 23; Indels 481; Gaps 4;

366 acgaacgggacacacccaatgagcgacatttcataataagaggtatgtctctcttc 445
DB 54044 ATGCATCTTTAAAGATGATCTTTATTTTCTTGTACCGAGAGGATATGTCCTCTTC 54103

CDS

54084..54514,54601..55709,56020..57468)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA94774.1"
/db_xref="GI:7630241"
/translation="MPNGLIRKLREELIDKLIATPPHPPOIPDVGAAAAADATAWICAPVAAANATCAPASFPVDPGVSPSCIRANPPPPNNRQTPRPVAAEPTDRIHPLIGCEWEDLLIHMGHGRGVYKADDEERKGRSARTEKLRGLGVKMLVDCGPIPIHGVGDGEMCVARVLLRMGSSNGRLLATLSLTAKNCLAMKLFPEKROTFELNPIEAAASGERSARARVDRAPPLRRGCRSRKRRPSCAPSLGSELYLASMCTKALIFVVAEKKIPFAGLIATFMGNLFLAIIILPARISGALIFHELIEVDIYCAVCSLIVAVIMETIVAVDLOARVYVTCMLARGSTGAVAPVLETTAPRODAAEACVDVDCCKLCKLKTSDNGTAPRLPAKONSTSDGCSITTFVRSITFVPASVSGVAVSPETOSGKNARSLPLPSVOEGNDSCNAPSGKNGAAEANDSPKMDLDQGNKONTDVANVSEDNTSDVGLAEPVQITWHLSEKAPSDGSPSTPEKLEVLVLRNDEBNKTEETLAEONCLTKDNPENSGKRRDOVAEONCLDKPRKPVSGKCEQICNEPCEVVLKSSKREKNAADCEEDMRKAKOOSKRRTAQAVSDALCKRRKPKKRVLLSEIINANOVRESSEYREKNAADCEEDMRKAPVPEVSMQIPVSNHTVGEIDGKSSKNKTRKRDVADVDESSLMNNKRRKRTGSHVHVHAPAGLSNKKVTPPTASTODENDENDENGCLDPMHKTCDVQGVSEISDORCSKKTACLSKGTHTSAATKTYGGESTRNCONTHVLSAEQCCQETENSVLSHAKSPAEHDQIMSLDHEQSLPKRRKQKLEVTREKQITDDIPMDVLELAKQKLEMLATEDCSDINRIGSTATTDDCVIAAKGSDYASSVEPTENSOOKSLASQKQLEQHLATLTDSPINPQNFSTQEQOQTHLRMEEMVITAAASPFLSHHDQYIEAEPTFHHGRDARKLAMEQKATRNPSAATCGAOFREMGVIOADVLTSHVSSSNVYASRODVLADIRVAYRAVNOYHARRPSTIATMEASLCLDRNAGOVVLYPREMSAPVLLDPMOESTLSPNTQSSRNQMSOLHNQYIAHNOYKGSTSYSGSNLNGKITPLETDELSDRLDHLNRPHPRVGLVLSLQKELIANWSENGCSTGSGTKLVSQVTSYSHQNNRERHEALNSGSAWNAQLQSVSSADPFLSARNSIAQSWTRGKKRVHPIDRFVRODICITKNKPADFTTISNDNEYMDYR

CDS

59368..59432,59585..59648,60333..60442,60620..60707,60835..60954))
/note="Ests AU082698(R0665), D23963(R0665) correspond to a region of the predicted gene.
Similar to Arabidopsis thaliana rac GTP binding protein Ara67. (AF079484)"
/codon_start=1
/protein_id="BAA94775.1"
/db_xref="GI:7630242"
/translation="MSSAAATRFKICVTVGDVAGTKCMLICYCNKPFDDYITPVE DNFANASVDSGVNLLGLMDPAGQEDYRLPLSLRGADVTLSSLSRSAYEWOK KAMEELRRFAPGVVLYGKTLIDREDAVYADHPASSLITTEGEEELKILIGAVAYI ECSSKELRNITKAVAPDAIKVYLOPRHMDVTRKQSSNNPVRVRYECGSACFA" join(67330..67530,67676..67816,69073..69234)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA94776.1"
/db_xref="GI:7630243"
/translation="MGSGSAAAPAPPMAMAGARWLEACVAPEDSGRRAFNALPIA GARSLASAGRAVCSILRTAELTDAEGMHGGLIAAADVDCAAIIGSEGIKIVSVH YDISYFSPARKLHEEVEDGRVVEQNGKATVATVELTRKDSGSELVATIGRQMSSTRPK DQASKSL join(69755..69967,70096..70236,70344..70514)
/note="Similar to Arabidopsis thaliana chromosome 1 BAC F19p19: unknown protein (AC000104)"
/codon_start=1
/protein_id="BAA94777.1"

QY	446	agaagaaaggaatccaaatctgctctcattctcgatttccatgaacgaataaatg	505
Db	54104	AGAAAGAGATCCAAAAATCTCTCTATCTCGGATTTTCCATGACCGAAGAAAATGTG	54163
QY	506	atgaacacaaagctagtccaagccatttctgtagaagatctgaacgaatggat	565
Db	54164	ATGAACACAAACCTAGTTCAACCCCAATTTCTGTAGCAAAAGTTTGACATGGGATGTCT	54223
QY	566	cgaaagtctctgataagttgaaaacttcagataatgaaacagacacaaatcttccg	625
Db	54224	CGAAGTGGTTGGATTAAGTTGAAGAACTCTCAGATTAATGGAACACACAGAACTCTTCCGG	54283
QY	626	caaaagcaagaatggccaagtgaatggtctcatcaatcattggttcggagactttgtgc	685
Db	54284	CAAAACAGAAATGGCCAAAGTGAATGGTGTCCATCATCAATTTGTGCGAACACTTTTGTGTC	54343
QY	686	ctgcgaagtgtgtgtcccaaaaagtgctccctcagaacacatcatctaaaggaagaatg	745
Db	54344	CTGCTAAGTGTGTGTTCCAAAAGTGTCCTCTAGCACACAAATCATCTCAAGGAGAAATG	54403
QY	746	ctgatagatcaaatcttcccaaaagatggtgcaaaagagcaatgaatcctcaaatgcaatgcgc	805
Db	54404	CTGATTAAGATCAACTCTTCCAAAAGATGTGCAGAAAGGCAATGACTCCAAATGAATGGCC	54463
QY	806	ctctcggcagaagaatggagctgctgtagagccaatactgatactcacaatgaa	854
Db	54464	CTTCTTGCGCAAGATGAGAGTGGTGTGAGGCGCAATATCATGATTCAACCAAGAAAGGATGTAG	54523
QY	855	-----	854
Db	54524	ATGTAGAGCCTTTCAAATTCCTTAAGTAGAATTTTATTAAAGTATAGAAATAAACTAATGT	54583
QY	855	-----agattgcaagggccagccccaataatgaatggtgcagcaaatg	899
Db	54584	TTGTGTGATTTTCTCGAGATTTTGCAGAGCGCAGCCAAAATTAATGATGTGGCAGCAATGT	54643
QY	900	ctctgaagcaaacactctgctgtgaatggtggcttaactgaagtccccaatcaatg	959
Db	54644	CTCTGAGGCAACACTCTTCTGTGATGTGTGGGCTTTTACTGGAAGTTTCCACAGATTACAGT	54703
QY	960	gaacatagaagtaaatggatggagatcaactcatcactccaatacttccgaagtgt	1019
Db	54704	GCACATAGAAAGTAATAGGTGCGAGATTAACCTCCATCCAAACTTCTGCAAGTGT	54763
QY	1020	cctccaagaataatgaatgtaaatgaaatgaaaacctgaagagactctgtgtgcgaagctg	1079
Db	54764	CCTCAAAAGAAATGAAGATGAAATGGAATACTGAAAGAGACTCTTGTGCTGAGCAGTG	54823
QY	1080	caattgaccaaagaatcctaaccceaatgtctgaaagaaagctgatacaggttgcgaaga	1139
Db	54824	CAATTTGACCAAAAGATCCTTAACCAATGCTGTGAAAGGAACGTAAGTGTGCTGAGCA	54883
QY	1140	gtgcgaattgacccaagaatccgaataacaaatgtcttgggcagaaatgtgaaacagatctcgaa	1199
Db	54884	GTGCAATTTTGACCAAAAGATCCGAACCAATGTCTGGGCGAAATGTGACAGATCTGCAAA	54943
QY	1200	tgaagcaatgtaagaagatgtgtctcctaaagaaagctccaatctaaaggaagaacgataa	1259
Db	54944	TGAGCCATGTGAAGAAAGTTGTTCTCAAAAGAAAGCTCCAAATCTTAAGAGGAACGATTA	55003
QY	1260	gaagtgtatgaagaagcaagcaagcaagcaagaagaacgacttgcgaagctgtatgttcaga	1319
Db	55004	GAAATGTGATGAAGAAAGCAGCAGCAGCAAGAAAGCAGCTGCCAGCGTGTGATTTCAAG	55063
QY	1320	tgcgaagcttgttcggagaagaagccaaaaaagtgtgcgtctcattcagaaattataatgc	1379
Db	55064	TGCAAAAGCTTTTGTGAGAAAGCAACCAAAAAGGTGCGGCTTCTATCAAGAAATTAATAATGC	55123
QY	1380	taacagagttgagaattctagaagaatgaacgaagtctaatctgtaaaatgtgcgtgcataccctg	1439
Db	55124	TAAACAGGTTGAGGATTTCTAGAAAGTACAGCAAGTTCACTGTGAAAGTGCCTGTATCCCTG	55183
QY	1440	tgaagatgaataaagatcacatcccggtcccgatggagaatgaacatgatatctctgttag	1499

Dp	55184	TTGAGGATGATRGAAGTACCATCCCGGTCGCGATGGAAGTAAAGATGATATCTCTGTGAG	55243
Qy	1500	caaccatacagctgggagaagaatgggttaaaataaagttaagaacaagacaaaacgaataa	1559
Dp	55244	CAACCATACAGTGGGACAAACATGGGTTAAATCAAGTAAGAACAAACAAACGCCAAATA	55303
Qy	1560	cttcgatgtttgtagatgatcatcatcaactaagaaactgctgtaactgtaaaaaaagaag	1619
Dp	55304	CTTCGATCTTGTAAATGATGATGATCATCATCTTAACTGGCTGAATGGAAAAAGAAAG	55363
Qy	1620	aactggaagtgatgatacaacacagttgtctatccaagctgggaattttagaacaacaaagt	1679
Dp	55364	AACTGGAGGTGTGATCCACACAGTTGGTCACTCCAGCTGGGAATTTAGCACACAAAAAGT	55423
Qy	1680	gaaccacacatcgagttactcaagatgaatgatgaataactgataactgtaaaatggtctgaac	1739
Dp	55424	GACACCCACTGCGATACTCATGACATGATGATGAGATGATGATACTGAAAAATGCTTGACAC	55483
Qy	1740	aaatatgataaagacagatgtctgttcagcatgatactcaagaatctccacaagatggtctc	1799
Dp	55484	AAATATGATTAAGACAGATGCTGTGCACACAGTATCGAAATCTCCACACAGAGTGTCTC	55543
Qy	1800	atcaaaagggaaaacacagcgggtttgagtaaggggaaaacacatctcagctgtagtaacaa	1859
Dp	55544	ATCMAAGGGAAACACAGGGGTTTAAGTAAAGGGGAAAAACATTCACACTGCTGATACAA	55603
Qy	1860	ataatggtgtgaagacacgaagaatggtgcagaacaaataatgatacagcgcaagaatga	1919
Dp	55604	ATATGTGTGTGAAGACACCAAGAAATGGTCAGAACATCATGTACTCAACGCGAAGATCA	55663
Qy	1920	atccagcagatggaacacgaacaaactctgtcttgatgaactcggcaa-----	1963
Dp	55664	ATCCACGATGGAACCGAAACCTCTGTCTGATGATCACTCGCCAAAGGTACGAATTTTGTG	55723
Qy	1964	-----	1963
Dp	55724	AATCATGAGGAATTTTGTCTTTTAAATTTGACTGAATCAACATTTATCTGATGAAGAA	55783
Qy	1964	-----	1963
Dp	55784	TAAATATTGGTCATTAACAATGTTAAAGAAATATGATACAATGTTTAAATATATGCTTCC	55843
Qy	1964	-----	1963
Dp	55844	ACTGTTCTTCTTACTTATCTTANGTTTGTGATCTCTTTTGTGTGGCGGCAATGTCATGT	55903
Qy	1964	-----	1963
Dp	55904	GTCGTGTGTCGTGTGTGTGGCGCGTGTGTGTGACAGTGTGGCGCAATATTCCTTTTTT	55963
Qy	1964	-----agttt	1969
Dp	55964	AGACTCATATTATAGTATTGTAAATGACATGACATTTTCTCTCATTTCTCATCTCAGGTTT	56023
Qy	1970	ctccagctgacatgatataccaatitaatgtctgaccttaatgacgaagatctaaccaaga	2029
Dp	56024	CTCCAGCTGACATGATATCCAAATTTATGTCGCTTCACTGACGAGAGCTTACCCAAAGA	56083
Qy	2030	agaaaagaagcaaaaacttgaagtgtcgtgtgaaaaaagacacatgatagatgacatcc	2089
Dp	56084	AGAAAAAACCAAAAACCTTTAACTGATGCTGTGAAAAACAGACCATGATGATGACATCC	56143
Qy	2090	ccaatgatattgttgaactgtcagctaaaaaaccagcatgaagagcagcttaatgactgaga	2149
Dp	56144	CCATGATATTCTTGAACTGCTAGCTAAAAACCGACATGAGAGCCACTTATGACGTGAGA	56203
Qy	2150	ctgattgttctgaacataccgattatcaatccaagaacacgtcgtatgatatgtgttaa	2209
Dp	56204	CTGATATTGTTCTGACATCAACCGTATTCATTCAAAGACAACCTGCTGATGATGTGTAA	56263
Qy	2210	tatgtagctggcaagatgattcaagattatgatacaagtgtgttgaacataatcccaac	2269

Dh	56264	TAGTAGCTGGCCAAAGATGTTTCAGATTATGATCAAGTGTGTTTGACACTAATTTCCCAAC	56323
Dh	2270	agaagctcttggaacatcccaagaatfacacagaaggagtattacaagggtcatcttbgcatctgacaa	2329
Dh	56324	AGAACTCTTGGCATTCGCCAAGTAACAGAAAGAGTATACAGGGGTCAATTTGGATTGGACCA	56383
Qy	2330	cacaaagatctcacatcctccgaacctttagctctactaagaagaacagacagacatacttgc	2389
Dh	56384	CACAAGACTCTCCACATCTCCAGAACTTTAGTCTACTCAGAACAGCAGACACAACTTTGGC	56443
Qy	2390	ggaatgaagaatbgtcaactatbgtcgtcaagctcacaccatatttcaacatcatgatgtc	2449
Dh	56444	GGATGGAAAGAAATGGTGCATCTATGCTCTCAAGCTACACACTAATTTTCACATCATGATGATC	56503
Qy	2450	agtatattgtctgaagaccaccaactgaaacatttggggcgctaaaggacgaagaagactaaagt	2509
Dh	56504	AGTATATTTGCTGAGACACCAACTGAACTTGGGGCGCTTAAGAGACGAAGAAAGCTAAAGT	56563
Qy	2510	gggagcaatttaagaagcactacaagaatactccagaagacaatactgtagtgcattta	2569
Dh	56564	GGAGCAATTTAAGGCCACTACAGAAATTCCTCAGCAGCAACATGGGGCTCAATTTTA	56623
Qy	2570	gaacttgatcccaagcagltgacttgacttctactatctgcatagggtcttccagaact	2629
Dh	56624	GACCTGGTATCCAAACAGTGTGACTTTGACTCTATGCTATGAGATCTTCCAGCAATT	56683
Qy	2630	atpcatctgcacaacagaaatctggccactgcgactgactgctgaaagacggttaacc	2689
Dh	56684	ATCATCTGCCACACAGTAATTTGGCCACTGGACCGCTATGCTGAAGAAAGCGGTTAACCC	56743
Qy	2690	aggctcactgcaagaanaatttccaaagcaaatagaacacatggaagcgagttaagtgtg	2749
Dh	56744	AGGTTCATGCAAGAAATTTTCCAAACACATAATGCAACCATGGAAGCGAGTAAGTTAGTG	56803
Qy	2750	atcgaggaaatctgtgcacagctagctctgtatcctaaagaaatcactgctgcagcgtc	2809
Dh	56804	ATCGAGCAATTCGTGGACAAAGTAAGTCTTATCTTAAGAAATTCATGCTTCGACGAGATC	56863
Qy	2810	tctctgaagatgatgatcactcaacattagcaaggttcccaactctgaactcttgaca	2869
Dh	56864	TTCTGAAATGAATGGATTCATCACTAACATTTGACAAAGCTTCCCAACTATGAACTTCTAGCA	56923
Qy	2870	ggaaccagatggagagctcacaactcacaattctctagatgacataatcgaatcaaaaggt	2929
Dh	56924	GGAAACAGATGGAGTCTCAACTTCATTAATTTCCAGTATGACATATATCAATGTAACAAAGAT	56983
Qy	2930	caaccagacacatcatatgacagtaacctgaatggaaaggtttccatttgcatctgaaagct	2989
Dh	56984	CAACCAGACATCATATATGACAGTAAGTTCGTAATGGAAGATTTCCATTGACATTCGAAGACT	57043
Qy	2990	tatacagagcatcagctgcatgcatctgcatgacttgaagaccttlaagccacaactctagagtgtg	3049
Dh	57044	TATCAGCGCATCAGCTGCAATGATCTGCACAGACTTTAGCCACACTCTTAAGTTGGTGTG	57103
Qy	3050	tgccttgagctctgctgcagaaaggaatctgcaaacctggtctcgagaagacttggcacacat	3109
Dh	57104	TGCTTGGCTCTCTCTGACAGAAAGAAATTTCCAACTGGTGTGGAAACTGTGGCACAAAT	57163
Qy	3110	ctgtgctataagtttaggagtggtgcacagagaataacatcgacatcagatgaaacagaagaaac	3169
Dh	57164	CTGTGTTATTAAGTTAGGATGTGCAACACAGAAATTAACATGCAATGATGAAACAGAAAGAAC	57223
Qy	3170	attctggaagccctggaatctctgaaatggttttcgaacaaatggaatgcatctgcaggttggtt	3229
Dh	57224	ATTTTGAAGCCCTGAAATCTTGGAATGTTTTCAGCAAAATGGAATGCAATTCAGTTGGGTT	57283
Qy	3230	ctgttagctcagctgcagaatttttatcagagaggaacagacatagctcacaactcttgacaa	3289
Dh	57284	CTGTTAGCTCCAGTCCACAAATTTTATCAGGGAAGAAACAGCATAGCTCAATCTTTGAGCA	57343
Qy	3290	ggggcgaaggttaaagtgtcatccctctgtagctggttgttgagacagagalaatctgtcataa	3349
Dh	57344	GAGGCAAGGGTAAATTGGTTCATCCCTTGATGCGTGTGAGACAGAGATATCTGTATAA	57403

OY	3350	ctaacaaagaccacgactgaatttctaacaatcagaaagataacgagatattgatatcc	3409
Db	57404	CTAACAAAGACCCGACTATTTTACTATAAATCAGTAACGATTACGATATATGATTTACC	57463
OY	3410	gtctgaagcagaagaagtgggtgtcatcaatcccgaaaccttacatcatacttaacctt	3469
Db	57464	GCTGAAGCAGAAGAAGGGTGTCATATAATTCCTGAACATTATTACAATCATATCAATTTCTTT	57523
OY	3470	atgggcgccaaaatgatcatcacctgtlaaaggaggagctttgctgatatgtctga-----	3521
Db	57524	ATGGGCCCAAATPATGCATATCGTAGAGAGAGGGCTTGTCTGGATCTGCTGAAGTAAGT	57583
OY	3522	-----	3521
Db	57584	TGAACTTTTCTCTTCAGCAAGTTTATTCAGTTTAAGAAAAAAGAAATGATTACTTATGTTAG	57643
OY	3522	-----aggtcttcgttaagtttgtagatgccccatttcigtatggagac	3565
Db	57644	CAGGATGCTTCTTCGACAGCTTCTTGTAAGTGTGTGATGCCCAATTTCTGGATGGGAAC	57703
OY	3566	ctgcacagacagltgaacaaggagcttgcagaagtgcagacatccgatttgttttcgaagtc	3625
Db	57704	CTGCACAGACAGTGAACAAGGCTTTCGAAGTGCAGCATCCGGTTTGTGTTTGGCAGTC	57763
OY	3626	caagaacagtcctccctgttacttggtagtgbtactacatactagiagcgctgtgttlacaa	3685
Db	57784	CAGAAGACGTCCTCCTGTACTTTGTATGTGATCATATAGTATGCGCTGTGTTGTACAA	57823
OY	3686	ggaggaatggtttaacctgttggaaaaaatgtctccccatttgttaattacataagaa	3745
Db	57824	GGAGAAAAGTGTTAACTCTGTTG-AAAAATGCTCCCCCAATTTGTATATTCATAAGAA	57882
OY	3746	ggattaagttgttggagactgtgtgagctgcagagcgagaaatggtttgtgcgggttaa	3805
Db	57883	GGTTATATGAGTGTGTGACCTGTGTGTGACTGACGGCGAAGATGGTTTGTGCGTTPAA	57942
OY	3806	ggttgaacagcactaactcgttatcaaatgtgtttaaactctgatatgtgttiacc	3865
Db	57943	GGTTGAACGACTACCTCTCGTATCAATGTGTTTAACTTCTGATTTGATGTGTATACC	58002
OY	3866	ttaacttgaagtcacacccgagaattiac	3896
Db	58003	TTACTCTTGAAGTCAACACCGAGAAATTATAC	58033
RESULT	3		
LOCUS	166494/c		
DEFINITION	Sequence 14 from patent US 5670367.		
ACCESSION	U66494		
VERSION	I66494.1 GI:2724471		
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 7218)		
JOURNAL	Dorner,F., Scheifflinger,F. and Falkner,F.Gunter.		
FEATURES	Recombinant fowlpox virus		
BASE COUNT	Patent: US 5670367-A 14 23-SEP-1997;		
ORIGIN	Location/Qualifiers		
	1..7218	/organism="unknown"	
	1944 a 1491 c 1486 g 1929 t 368 others		

Query Match	1.6%	Score 62.8;	DB 6;	Length 7218;
Best Local Similarity	7.2%;	Pred. No. 1.2e-05;		
Matches	31;	Conservative 227;	Mismatches 174;	Indels 0;
				Gaps 0;
QY	1182	atgttagcagatctcgaatgagccatctggaagaattcttccaagaagctccaatc	1241	

Dn	1497	ATGAGGCATCATCTGAATTACATTCATCATCAAGTACTTAAGAAGATAGAAGAAATTGG	1438
Oy	1242	taagaagaaacgcgataagaagtgtgaagaagaacagcagcagcaagaagacactgc	1301
Dn	1437	TACRRR	1378
Oy	1302	cgaagctgatcttcagatcgaaagcttctgcgagaagccaagaagctgcctct	1361
Dn	1377	RRR	1318
Oy	1362	atcgaaatlaaatgctaaccsgtgtgagatcttagaagtgacgaagtlcatctga	1421
Dn	1317	RRR	1258
Oy	1422	aatgycgctgattccctgttgaggatgaagaagtaacctcccgtlcccgaggaagta	1481
Dn	1257	RRR	1198
Oy	1482	catgatatcttcgttagcaaccatacagctggagaaagtgtgtaaataaagtaaga	1541
Dn	1197	RRR	1138
Oy	1542	caagaacaaagcgaataactctgtatgtgtgtagatgatactcaactatgaactgc	1601
Dn	1137	RRR	1078
Oy	1602	gaatcgaaaaa 1613	
Dn	1077	RRRRRRRRRRRA 1066	
RESULT	4		
AC026135		141041 bp DNA linear HTG-05-APR-2000	
LOCUS		Homo sapiens chromosome 3 clone RP11-78022 map 3, WORKING DRAFT	
DEFINITION		SEQUENCE, 12 unordered pieces.	
ACCESSION		AC026135	
VERSION		AC026135.2 GI:7417828	
KEYWORDS		HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
TITLE		1 (bases 1 to 141041)	
JOURNAL		Birren,B., Linton,L., Nusbaum,C. and Lander,E.	
REFERENCE		Homo sapiens chromosome 3, clone RP11-78022	
AUTHORS		Unpublished	
		2 (bases 1 to 141041)	
		Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,	
		Anderson,S., Baldwin,U., Barra,N., Bastien,V., Beeda,F.,	
		Boguslavsky,I., Boukhalter,B., Brown,A., Burtelt,G.,	
		Campiolino,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,	
		Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,	
		Dodges,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,	
		Galaen,J., Gardina,S., Glnde,S., Goylette,M., Graham,L.,	
		Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,	
		Howland,O.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,	
		Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,	
		Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,	
		McCarthy,M., McEwan,P., McGurv,T., Miranda,C., Menga,V., Morrow,J.,	
		Meldrum,J., Menes,L., Mihov,T., Miranda,C., Mlangi,V., Morris,J.,	
		Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,	
		O'Neill,D., Oliver,T.M., Oliver,C., Peterson,K., Pierre,N.,	
		Pisanti,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,	
		Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,	
		Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,	
		Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,D.,	
		Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,	
		Young,G., Zainoun,J., Zimmer,A. and Zody,M.	
TITLE		Direct Submission	
JOURNAL		Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome	
		Research, 320 Charles Street, Cambridge, MA 02141, USA	
COMMENT		On Apr 5, 2000 this sequence version replaced gi:7264205.	

Dugan-Rocha, S., Durbin, K.J., Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, N., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C., Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W., Loulsegad, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogund, M., Okwuon, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gdbbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Unpublished
2 (bases 1 to 141095)
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCGD
Center clone name: RP11-78022
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 133822 bases at least Q40
Consensus quality: 137507 bases at least Q30
Consensus quality: 138991 bases at least Q20
Estimated insert size: 139059; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-IP estimation
Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraft.data.html>).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
*
* 1 34364: contig of 34364 bp in length
* 34365 34464: gap of unknown length
* 34465 69010: contig of 34546 bp in length
* 69011 69110: gap of unknown length
* 69111 84563: contig of 15453 bp in length
* 84563 84664: gap of unknown length
* 84664 99240: contig of 14577 bp in length
* 99240 99341: gap of unknown length
* 99341 109623: contig of 10283 bp in length

* 109624 109723: gap of unknown length
* 109724 123170: contig of 13447 bp in length
* 123171 123270: gap of unknown length
* 123271 129353: contig of 6083 bp in length
* 129354 129453: gap of unknown length
* 129454 134225: contig of 4772 bp in length
* 134226 134325: gap of unknown length
* 134326 138799: contig of 4374 bp in length
* 138799 138700 141095: contig of 2296 bp in length.
* 138800
Location/Qualifiers
1. 141095
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q"
/clone="RP11-78022"

BASE COUNT 45088 a 25322 c 25884 g 43887 t 914 others
ORIGIN
Query Match 1.2%; Score 48.2; DB 2; Length 141095;
Best Local Similarity 46.5%; Pred. No. 0.17;
Matches 155; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 1418 gTgaataatgcgcgtgacccctgtgagatagaagatccctccgcgcgagtgagag 1477
DB 11830 GAGATCATGCCATTACCTCCAGCCTGTCACAGACGACAGATTCATCTCAAAAAA 11771
QY 1478 taagcatgatactctctgtgaacacatagatgaggaagatgggttaaatcaagta 1537
DB 11770 AAAAAAGTAATTTGGCATTTAGGAAACATTTGTAAGAGAGGGGTAAATGTTAACAA 11711
QY 1538 agaacagaacaaacgcaaatctctgtatgtgtgatgataatgatacacttaagaact 1597
DB 11710 ATAGCAAAATCAAAACCAACCACTTATGACTTAAAAAAGATTTGAAAC 11651
QY 1598 ggcctgataagaaaagaagaacgtgagtgatgacacacagtgctcaccgctg 1657
DB 11650 CTGATGGAGAAAATCTGTAAACCTGATTAATTAATTCATTTAATGTTCTCGAAGATA 11591
QY 1658 ggaatttgacaacaaaagatgacacccacgtcgagtaactagcatgtatgaagaa 1717
DB 11590 GACATTTGAAGCAACAAACAGAAAATTAATTAATTAATTAATTAATTAATTCG 11531
QY 1718 ataccgaaaatgctcttgacacaaatagcata 1750
DB 11530 AAAAACAAAATGATTTCTATTGAAATTCATA 11498

RESULT 6
LOCUS AL591122 189409 bp DNA linear HTG 18-JAN-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-415K20. *** SEQUENCING IN PROGRESS ***; 3 unordered pieces.
ACCESSION AL591122
VERSION AL591122.13 GI:16904444
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Martin, S.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
TITLE
AUTHORS
JOURNAL

Homo sapiens chromosome 11, clone RP1-220G17
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 114980)

REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,I., Boukhalter,B., Brown,A.,
Camarata,J., Campioano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Glade,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Labocque,K., Lamarez,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P.,
Margulis,N., Matthews,C., McCarthy,M., McMan,P., McKernan,K.,
McPheters,R., Meldrim,J., Menes,L., Mihov,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Punkhang,P., Pierre,N., Pollara,V., Raymond,C., Retter,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schnapack,R., Seaman,S., Severy,P.,
Sognez,C., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Travers,M., Trevis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo/A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted
Direct Submission
Submitted (28-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 24, 2001 this sequence version replaced g1:12584323.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: LI2108
Center clone name: 220.G.17
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; Version 0.960731
Consensus quality: 114348 bases at least Q40
Consensus quality: 114583 bases at least Q30
Consensus quality: 114709 bases at least Q20
Insert size: 115000; agarose-fp
Insert size: 114780; sum-of-contigs
Quality coverage: 10.3 in Q20 bases; agarose-fp
Quality coverage: 10.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 3 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 5011: contig of 5011 bp in length
* 5012 5111: gap of 100 bp
* 5112 73744: contig of 68633 bp in length
* 73745 73844: gap of 100 bp
* 73845 114980: contig of 41136 bp in length.
Location/Qualifiers
1. 114980
/organism="Homo sapiens"

RESULT 12
AC022048
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 158392)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-15F4
Unpublished
2 (bases 1 to 158392)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beta,F.,
Boguslavsky,I., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dominko,M., Doyle,M., Fenesler,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McMan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrim,J., Menes,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Trigilio,J., Travers,M., Trevis,N., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP1-220G17"
/clone_lib="RPc1 Human PAC library 1"
1. 5011
/note="assembly-fragment
vector_side:left"
clone_end:SP6
5112..73744
/note="assembly-fragment"
73845..114980
/note="assembly-fragment
clone_end:T7
vector_side:right"

BASE COUNT 37057 a 20292 c 20482 g 36949 t 200 others
ORIGIN

Query Match 1.2%; Score 45.2; DB 2; Length 114980;
Best Local Similarity 52.1%; Pred. No. 1.1;
Matches 101; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 2799 tggagcgacatcttggaagtgtgatcatcaacaattgaaagttccccaactatg 2858
Db 74953 TACATCTCTTTTCCCAATCATGCATCAACATCTATTATTTTTGGATTTAAATTAGG 75012
Qy 2859 aaacttcagcagaaccagatgagttcctaactcatacttcagtatgcataatca 2918
Db 75013 CCATTCCTTGACGAGTAGTAAGCGTGCTCATGTGCTGAATTCATTCCTCGATTA 75072
Qy 2919 gtacaaggatcacaccgacatcatatgtagtgcagtaacctgaatggaagaattccattgac 2978
Db 75073 TTACTGATGTGACCATTTTCTTCATATGTTCTGCTGCTGTTTGATATCTTCCTTTGAG 75132
Qy 2979 attcgaagactat 2992
Db 75133 AATTGTCTATTAT 75146

TITLE
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 2, 2001 this sequence version replaced gi:100453377.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3455
 Center clone name: 15_F_4

----- Summary Statistics

Sequencing vector: M13; M7815; 36% of reads
 Sequencing vector: Plasmid; n/a; 64% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 158197 bases at least Q40
 Consensus quality: 158232 bases at least Q30
 Consensus quality: 158245 bases at least Q20
 Insert size: 16300; agarose-fp
 Insert size: 158292; sum-of-contigs

Quality coverage: 13.8 in Q20 bases; agarose-fp
 Quality coverage: 14.2 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 39274: contig of 39274 bp in length
 * 39275 39374: gap of 100 bp
 * 39375 158392: contig of 119018 bp in length.

FEATURES

source

1. 158392

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Rp11-15F4"

/clone_1fb="RPI-11 Human Male BAC"

misc_feature

1. 39274

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

39375. 158392

/note="assembly_fragment"

clone_end:T7

vector_side:right"

BASE COUNT 48372 a 27087 c 29475 g 53358 t 100 others
 ORIGIN

Query Match 1.2%; Score 45.2; DB 2; Length 158392;
 Best Local Similarity 52.1%; Pred. No. 1.1;
 Matches 101; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

OY 2799 tgcgagcatcttcgtgagaatgatccatcaatagcaagcttcccaactatg 2858
 DB 89005 TACATCTCTTTTGGCCACATCCATCAATCATATATTTTGAATTTTAAATTATGG 89064
 OY 2859 aactctagcaggaacacagatgagctcaactcataatcttcagatgacataatca 2918
 DB 89065 CCAATCTTGACAGAGTAAGGTGTCATGTCATGTTCTTAATTCGATTTCCCATATNA 89124
 OY 2919 gtacaaaggatcaaccagacatcatatgagcagtaacctgaatggaagaatccattgac 2978
 DB 89125 TTACTGATGTTGAGCATTTTCTTCAATATGTTGCTGCGGTTTGAATATCTTCCTTTGAG 89184

OY 2979 attcgaagacttat 2992
 DB 89185 AATTGCTCTATTAT 89198

RESULT 13

SH1U145A/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

source

1. 3400

/organism="Pseudorabies virus"

/strain="Indiana-Punkhauser"

/db_xref="taxon:10345"

/note="The virus was passed in Madin-Darby bovine kidney cells"

217. 2718

/gene="UL5"

217. 2718

/gene="UL5"

/note="UL5"

/note="putative"

/codon_start=1

/product="helicase"

/protein_id="AAA50986.1"

/db_xref="GI:456681"

/translation="MAAAGAGPATNATNTNPSMHCVPPIVRELAATPAEAP

PPLAFKSLAANDPNVDLEALEPRAATLISGNAGSGSTCIOTINEMDCYTIGSTR

VAAQNVVAKLSAAYSSRYVNTIPEDEGRGNHVOQLGRYSIACPTSPPTVARELOKRD

LVTWVEVLQDISRYVLAGSHEFARLALERTGAAHLLAFACGSLPAPARSNVIV

IDEAGILGRHLITAVVYCMWLIINAAYDTTOYAAARPVLCVGSPTQDSLESREHA

ROLCHVRASENLITVLTINRALREYTDLSRNAIFINNRKCOEYFEGELMALEGLP

LTDEHLRLVDSFVPEAYTINPANLQGTPLXSSHREVSAYMSRLHAHLKYAGDAOPV

VFTLPATYITVIRNAEDRYREATQPHRLTLDRHLAANKRRITNYSORDDAALACEA

RAQOGVILARCENVYVINSOVAVTRRLKLVIGSGTEAPAAVLRDAFVHAQGS

EYTRFLLSLFSGMIATYNELOPGLAPAVTAARLAATVTAALVPEHEDPFGS

AAAPAPAGPGGADDDDLFALSENMLDMLYCHYDFARETSEVYAOFLMKTLF

ADRYAALSELFGAPAPAPFETHVDVSVRGCEVVGGLRGLISTALQTSYTLVYG

THATAPAPAEELARKKLKLAAGTAEELASLDTPRVVLRDQSGPLSTINVLSPVEELD

LELDNATTVDTGISKLAMTARSQGLSLERAAVCFTRANLMMNSVYVAMSVISRF

LRMNNNPRLREPERDNGISEHILALRDGAHVIY"

2777. 3214

/gene="UL4"

2777. 3214

/gene="UL4"

/note="UL4"

/function="unknown"

/note="putative"

/codon_start=1

/protein_id="AAA50987.1"

/db_xref="GI:431466"

/translation="MTALVAYSFYRIKLPDGMADPGCGOTVCEVGRGVNATDQCTR

CDALAPRGVTTIOHGVLTIVLVADGEPKCSYFATFAPRAAREGALVMPFSWSCAERS

RRLRGPAGGLATVLAERALHNTTITVAYRPDVLRLREARILE"

BASE COUNT 502 a 1241 c 1160 g 497 t

ORIGIN


```

* 25744 25843: gap of 100 bp
* 25844 26559: contig of 716 bp in length
* 26560 26659: gap of 100 bp
* 26660 27349: contig of 690 bp in length
* 27350 27449: gap of 100 bp
* 27450 28145: contig of 696 bp in length
* 28146 28345: gap of 100 bp
* 28346 28956: contig of 711 bp in length
* 28957 29056: gap of 100 bp
* 29057 29789: contig of 733 bp in length
* 29790 29889: gap of 100 bp
* 29890 30611: contig of 722 bp in length
* 30612 30711: gap of 100 bp
* 30712 31413: contig of 702 bp in length
* 31414 31513: gap of 100 bp
* 31514 32239: contig of 726 bp in length
* 32240 32339: gap of 100 bp
* 32340 33039: contig of 700 bp in length
* 33040 33139: gap of 100 bp
* 33140 33833: contig of 694 bp in length
* 33834 33933: gap of 100 bp
* 33934 34638: contig of 705 bp in length
* 34639 34738: gap of 100 bp
* 34739 35425: contig of 687 bp in length
* 35426 35525: gap of 100 bp
* 35526 36220: contig of 695 bp in length
* 36221 36320: gap of 100 bp
* 36321 37019: contig of 699 bp in length
* 37020 37119: gap of 100 bp
* 37120 37831: contig of 712 bp in length
* 37832 37931: gap of 100 bp
* 37932 38591: contig of 660 bp in length
* 38592 38691: gap of 100 bp
* 38692 39379: contig of 688 bp in length
* 39380 39479: gap of 100 bp
* 39480 40189: contig of 710 bp in length
* 40190 40289: gap of 100 bp
* 40290 41022: contig of 733 bp in length
* 41023 41122: gap of 100 bp
* 41123 41813: contig of 691 bp in length
* 41814 41913: gap of 100 bp
* 41914 42623: contig of 710 bp in length
* 42624 42723: gap of 100 bp
* 42724 43442: contig of 719 bp in length
* 43443 43542: gap of 100 bp
* 43543 44274: contig of 732 bp in length
* 44275 44374: gap of 100 bp
* 44375 45094: contig of 720 bp in length
* 45095 45194: gap of 100 bp
* 45195 45925: contig of 731 bp in length
* 45926 46025: gap of 100 bp
* 46026 46761: contig of 736 bp in length
* 46762 46861: gap of 100 bp
* 46862 47535: contig of 674 bp in length
* 47536 47635: gap of 100 bp
* 47636 48373: contig of 738 bp in length
* 48374 48473: gap of 100 bp
* 48474 49204: contig of 731 bp in length
* 49205 49304: gap of 100 bp
* 49305 50024: contig of 720 bp in length
* 50025 50124: gap of 100 bp
* 50125 50806: contig of 682 bp in length
* 50807 50906: gap of 100 bp
* 50907 51608: contig of 702 bp in length
* 51609 51708: gap of 100 bp
* 51709 52417: contig of 709 bp in length
* 52418 52517: gap of 100 bp
* 52519 53229: contig of 712 bp in length
* 53230 53329: gap of 100 bp
* 53330 54064: contig of 735 bp in length
* 54065 54164: gap of 100 bp
* 54165 54877: contig of 713 bp in length
* 54878 54977: gap of 100 bp

```

```

* 54978 55714: contig of 737 bp in length
* 55715 55814: gap of 100 bp

Query Match      1.1%; Score 44.2; DB 2; Length 65478;
Best Local Similarity 48.2%; Pred. No. 1.9;
Matches 124; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 1041 aaatggaanaacgcgaagagactctgttctgtgcgacgtgcaattgacccaagatccaa 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20758 AAAGAGAAATTAAGAAAGATTAAGACCTTCTAGAAATTCAGAAAGAAATGACATACCAT 20699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1101 cccaatgctggaagaaggaacgtgatccaggttgcgcagtcgaattgacccaagatcc 1160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20698 ACCAATTATATGGACACACAAAGTAAAGCACTACTAAGACGATATTCATAGCATTAAGGCC 20639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1161 gaaaccagtgctcgtgcagaatctgtgacagatctgcaatgagccatggaagaagttgt 1220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20638 TTCATTAAGAAATTTGGAAACATACATCAAGATTTAAGAAATACCTTGAAGAAAGAA 20579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1221 tctcaaaagaagctccaatctaagaaggaagcgataagaagtgtgtgaagaagcgca 1280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20578 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 20519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1281 gcaacagcaagaagcgca 1297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20518 GAAGAAAGAAAGAAAGGGA 20502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15

```

AL133232
LOCUS      Human DNA sequence from clone RP5-897D18 on chromosome 20. Contains
DEFINITION part of a gene for a putative novel protein, EST, STS and GSS,
            complete sequence.
ACCESSION  AL133232
VERSION    AL133232.15 GI:8649008
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      1 (bases 1 to 95097)
JOURNAL    Direct Submission

```

```

COMMENT    Submitted (14-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Jun 21, 2000 this sequence version replaced gi:8574104.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/RGP/Chr20

```

```

            This sequence is the entire insert of clone RP5-897D18. The true
            left end of clone RP11-4603 is at 76409 in this sequence. The true
            right end of clone RP4-539E24 is at 40325 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP5-897D18 is from

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 16:58:13 ; Search time 563.69 Seconds
(without alignments)
11866.626 Million cell updates/sec

Title: US-09-828-068-1

Perfect score: 3896
Sequence: 1 cgcgcgtctgcgagcaaacg.....gtcaacaccgagaatttacc 3896

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq-032802:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	43	1.1	1788	23	ABLO6815
C 2	43	1.1	4773	23	ABLO6814
C 3	42.8	1.1	574	21	AA44195
4	41.4	1.1	246	23	AA590679
5	41.4	1.1	2456	23	AA575925
C 6	41.4	1.1	2848	22	AAH54196
7	40.8	1.0	2167	21	AA532967
8	40.4	1.0	686	21	AAA09318
9	40.2	1.0	15518	24	ABL34172

10	40.2	1.0	15518	24	ABL34624	Human metastasis a
C 11	39.8	1.0	15122	24	ABL32775	Human immune syste
C 12	39.6	1.0	3257	22	AAH54703	S. epidermidis gen
C 13	39.4	1.0	843	23	AA592080	DNA encoding novel
C 14	39.2	1.0	2167	21	AA538223	Human S184L mutant
15	39.2	1.0	2522	19	AAV63196	CDNA from clone fp
C 16	39.2	1.0	6065	24	ABL32504	Human immune syste
C 17	39	1.0	833	22	AA194740	Human neuroblastom
18	38.8	1.0	567	21	AA295550	HIV codon altered
19	38.8	1.0	14041	22	AAH48024	Internal control B
20	38.6	1.0	261	22	ABA49199	Human breast cell
C 21	38.6	1.0	261	22	ABA67114	Human foetal liver
C 22	38.6	1.0	261	22	ABA75585	Human foetal liver
C 23	38.6	1.0	261	22	ABA34204	Probe #12670 for g
C 24	38.6	1.0	261	22	ABA40197	Probe #18663 for g
C 25	38.6	1.0	261	22	AAK15554	Human brain expres
C 26	38.6	1.0	261	22	AAK24165	Human brain expres
C 27	38.6	1.0	261	22	AAK41287	Human bone marrow
C 28	38.6	1.0	261	22	AAK50216	Human bone marrow
C 29	38.6	1.0	261	22	AA122038	Probe #11971 for g
C 30	38.6	1.0	261	22	AA127301	Probe #17234 for g
C 31	38.6	1.0	261	22	AA147350	Probe #16016 used
C 32	38.6	1.0	261	22	AA156167	Probe #24853 used
C 33	38.6	1.0	261	22	AA107733	Probe #7724 used t
34	38.6	1.0	448	22	ABA52290	Human foetal liver
35	38.6	1.0	448	22	ABA22091	Probe #557 for gen
36	38.6	1.0	448	22	AAK00565	Human brain expres
37	38.6	1.0	448	22	AAK26014	Human bone marrow
38	38.6	1.0	448	22	AA110643	Probe #576 for gen
39	38.6	1.0	448	22	AA131897	Probe #583 used to
40	38.6	1.0	448	22	AA100573	Probe #564 used to
C 41	38.6	1.0	513	22	ABA63135	Human foetal liver
C 42	38.6	1.0	513	22	ABA30385	Probe #8851 for ge
C 43	38.6	1.0	513	22	AAK11566	Human brain expres
C 44	38.6	1.0	513	22	AAK37335	Human bone marrow
C 45	38.6	1.0	513	22	AA118167	Probe #8100 for ge

ALIGNMENTS

RESULT 1	ABL06815/c	ABL06815 standard; cDNA; 1788 bp.
XX	AC	ABL06815;
XX	26-MAR-2002	(first entry)
XX	Drosophila melanogaster	expressed polynucleotide seq ID NO 14927.
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;
XX	KW	pharmaceutical; gene; ss.
OS	Drosophila melanogaster.	
XX	PN	W0200171042-A2.
XX	PD	27-SEP-2001.
XX	PE	23-MAR-2001; 2001WO-US09231.
XX	PR	23-MAR-2000; 2000US-191637P.
XX	PR	11-JUL-2000; 2000US-0614150.
PA	(PEKE)	PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX	WPI; 2001-656860/75.	
DR	P-PSDB; ABB62712.	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more	

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.1%; Score 42.8; DB 21; Length 574;
Best Local Similarity 57.5%; Pred. No. 0.095;
Matches 77; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
```

```
QY 1251 gacggataagaagtgtgataagaagcagcacagcaagaacgcacgtgcccgagctga 1310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 gaagttgaagaagttgataagaagaagaagaagaagaagaagaagtgtgattgtga 181
QY 1311 ttttcgagtcgaagctgtgtcgagaagaagccaaaaggtgcgctctatcagaat 1370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 182 tgaatgagtagtggtgaattcttcgtaaaaaaagcaagaagatgaaattcaacatcagt 241
QY 1371 tataatgtctaacc 1384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 242 ttccagacaataatc 255
```

```
RESULT 4
ID AAS90679 standard; cDNA; 246 BP.
AC AAS90679;
XX
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #26483.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
PR
```

```
XX
PA (HXSE-) HXSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABG26492.
PT
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 26483; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 246 BP; 126 A; 20 C; 77 G; 23 T; 0 other;
```

```
Query Match 1.1%; Score 41.4; DB 23; Length 246;
Best Local Similarity 60.0%; Pred. No. 0.14;
Matches 69; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1178 agaattgagcagatctgcacatgacatgtgagaagttgttccaaagaagctoca 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71 aggaactgtgaagaagaagaagaagaagactgcgaagaagaagaagaagaagaagaaga 130
QY 1238 aatctaagaagaagcagatagaagattgtatgaagaagcagcacagcaaga 1292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 131 agaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 185
```

```
RESULT 5
ID AAS75925 standard; cDNA; 2456 BP.
AC AAS75925;
XX
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #11729.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX
```

XX	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
PI	Drmanc RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
DR	P-PSDB; ABG11738.
XX	
XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 1; SEQ ID NO 11729; 103bp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and genome mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SO	Sequence 2456 BP: 656 A; 575 C; 654 G; 571 T; 0 other;
Query Match 1.1%; Score 41.4; DB 23; Length 2456;	
Best Local Similarity 60.0%; Pred. NO. 0.6;	
Matches 69; Conservative 0; Mismatches 46; Indels 0; Gaps 0;	
OY	1178 agaatgtcagagatctgcacatgagccatgagagaagtgttctccaaagaagtcaca 1237
DB	2281 aggactgtgaaagaagaagaagaagactgcgaaagaagaagaagaagaagaaga 2340
OY	1238 aatcagaagagagacgcgataagaagtgtgatgaagaagagacacacaagaagaa 1292
DB	2341 agaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaa 2395
RESULT 6	
AAH54196/C	
AAH54196 standard; DNA; 2848 BP.	
AC	AAH54196;
XX	
DT	03-SEP-2001 (first entry)
XX	
DE	S. epidermidis genomic polynucleotide sequence SEQ ID NO:3560.
XX	
KW	Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX	vaccination; endocarditis; ds.
OS	Staphylococcus epidermidis.
XX	
PN	WO200134809-A2.
XX	

PD		17-MAY-2001.
XX		
PF	09-NOV-2000;	2000MO-US30782.
XX		
PR	09-NOV-1999;	99US-0164258.
XX		
PA	(GLAX)	GLAXO GROUP LTD.
XX		
PI	Kimmerly WJ;	
DR	WPI; 2001-316495/33.	
XX		
PT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,	
PR	useful for vaccinating against infections, e.g. endocarditis -	
PS	Claim 8; Page 1144-1145; 218Bpp; English.	
XX		
CC	AAH5304 to AAH53970 represent nucleic acids (I) encoding polypeptides	
CC	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.	
CC	(I) and (II) can have antibacterial activity and therefore can be used	
CC	in vaccination. The nucleic acids (I) may be used to produce the	
CC	S. epidermidis polypeptides (II) via the production of vectors	
CC	containing them which are used to produce hosts cells which express the	
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be	
CC	used to vaccinate subjects and to raise antibodies against the bacteria.	
CC	The polypeptides may also be used to assay for other inhibitors of their	
CC	activity and therefore identify compounds that may be used for the	
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to	
CC	AAH55080 represent specifically claimed S. epidermidis genomic DNA	
CC	polynucleotide sequences from the present invention. AAH55091 to	
CC	AAH55098 represent oligonucleotide sequences and primers which are used	
CC	in the exemplification of the present invention.	
CC	N.B. The present invention specifically claims all the polynucleotide	
CC	sequences given in the sequence listing of the present specification,	
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even	
CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,	
CC	no sequences are present for SEQ ID NO:4455 to 4464.	
SQ	Sequence 2848 BP; 876 A; 501 C; 442 G; 1029 T; 0 other:	
Query Match	1.1%;	Score 41.4; DB 22; Length 2848;
Best Local Similarity	49.8%;	Pred. No. 0.66;
Matches 105; Conservative	0;	Mismatches 106; Indels 0; Gaps 0;
OY	713	ccctcgacacacatcatcctcaagggaagaaagcgcgatgataccaactcttcacaaagtgtg 772
DG	2021	crrtaagtgtaaaaagaacaatcgaaacgtaaaanacagagccggcgatttagctcaattatgag 1962
OY	773	tgcagaagagcaatgactcctcaaatgcaatgcccctcttgagcaagaatgtagtctgtgag 832
DG	1961	aagagacatttaatttgacatcattgattgattgacctttcttgaanaattgaacttaagtag 1902
OY	833	ccaatactatccaccatgaagaatttcgaaggscacgcccaaaaaattagatgtgycag 892
DG	1901	tcatttactgaacaaacaatttatgaaagaattgacgcctgaattatattgatatgaaagaa 1842
OY	893	caaattgctctgaggacaacactcttgttga 923
DG	1841	atcttaatttaattgaacaaacacatctgcttta 1811
RESULT	7	
ID	AAZ32967	
XX	AAZ32967 standard; DNA; 2167 BP.	
AC	AAZ32967;	
DT	09-FEB-2000 (first entry)	
DE	Human wild-type protein kinase MKK4 gene fragment E.	
XX		
XX	MKK4; mitogen activated protein kinase; MAPK; MAPK pathway; mutation;	

KM somatic; signal transduction; apoptosis; stress; cytokine; induction;
 KM phosphorylation; Jun kinase; JNK; p38; tumour; suppressor;
 KM loss of heterozygosity; LOH; cancer; detection; diagnosis; prognosis;
 KM breast cancer; pancreatic cancer; colorectal cancer; testicular cancer;
 KM drug screening; gene therapy; protein replacement therapy; mimetic; ds.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Intron 1..247
 FT /*tag= a
 FT exon 248..367
 FT /*tag= b
 FT /note= "MKK4 exon E"
 FT Intron 368..2167
 FT /*tag= c
 XX
 PN US5989885-A.
 XX
 PD 23-NOV-1999.
 XX
 PF 13-JUN-1997; 97US-0874186.
 XX
 PR 10-JAN-1997; 97US-0782482.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 XX
 PI Skolnick MH, Perry WL, Tavtigian SV, Teng DH-;
 XX
 DR WPI; 2000-022786/02.
 XX
 PT Polynucleotides comprising all or a portion of the tumor suppressor
 PT gene MKK4 locus are useful for diagnosis, prognosis and therapy of
 PT human cancers -
 XX
 PS Claim 1; Columns 71-74; 63pp; English.
 XX
 CC This sequence represents a fragment of the gene encoding a human wild-
 CC type protein kinase, MKK4, which includes exon E. The gene comprises 11
 CC coding exons. MKK4 (also known as JNK1 and SEK1) may be involved in a
 CC MAPK (mitogen-activated protein kinase) pathway for the signal
 CC transduction of cytokine-induced and stress-induced apoptosis. MKK4 is
 CC also involved in suppressing a variety of tumours. MKK4 is a dual
 CC specific kinase that activates Jun kinases (JNKs) and p38 (a MAPK) but
 CC not extracellular signal-regulated kinases (ERKs) which are a subgroup
 CC of MAPKs. The JNK and p38 MAPKs are activated via dual phosphorylation
 CC on threonine and tyrosine and then go on to activate proteins further
 CC downstream in signal transduction pathways. Tumour suppressor genes such
 CC as MKK4 are deleted at high frequency in certain tumour types. The
 CC deletions often involve loss of a single allele, which is known as loss
 CC of heterozygosity (LOH), and the remaining allele is presumed to be
 CC non-functional, either because of a pre-existing inherited mutation, or
 CC because of a secondary sporadic mutation. Alternatively, the deletion
 CC may involve homozygous deletion of both alleles. LOH events commonly
 CC involve deletions spanning many megabases of DNA, while homozygous
 CC deletions are relatively small in size, probably due to the proximity of
 CC essential genes. Sequences derived from the MKK4 gene can be used to
 CC detect a portion of the MKK4 locus or its expression product in a tissue
 CC sample for the diagnosis and prognosis of human cancer, and can
 CC also be used to diagnose of a predisposition to breast, pancreatic,
 CC colorectal and testicular cancers, as specific MKK4 mutations have
 CC been found in cell lines derived from such tumours. MKK4 oligonucleotides
 CC are useful for the detection of the nucleotide sequence of a particular
 CC MKK4 allele via PCR, and can be used as probes to detect point mutations,
 CC PCR amplification products and mismatches between the MKK4 gene or mRNA.
 CC MKK4 proteins can be used for screening of drugs which can restore MKK4
 CC gene product function for cancer therapy. MKK4 gene therapy, protein
 CC replacement therapy and protein mimetics that reconstitute the function
 CC of the MKK4 protein may be used for therapy of human cancers which result
 CC from a mutation in the MKK4 gene.
 XX
 S0 Sequence 2167 BP; 716 A; 334 C; 347 G; 768 T; 2 other;

Query Match 1.0%; Score 40.8; DB 21; Length 2167;
 Best Local Similarity 54.7%; Pred. No. 0.83; Indels 0; Gaps 0;
 Matches 81; Conservative 0; Mismatches 67;
 QY 3610 ttgttttgcacgacgaacgctcctctgtacttctgtactgttactactactagt 3669
 D 171 tatgtattccatttaagtaagcaggtgatatattagaatgataagaataacaga 230
 QY 3670 ggcgtctgtttacaaagagaatgttaacctgtttgaaaaaatgtctcccatitt 3729
 D 231 tatgttattttaaaggttactgttgcattgtatgaacctatgtctactccctgtt 290
 QY 3730 gtaattaccataagagattatagtt 3757
 D 291 ataagtttacaataatgatatagttgt 318
 RESULT 8
 AAA09318
 ID AAA09318 standard; DNA; 686 BP.
 XX
 AC AAA09318;
 XX
 DT 10-AUG-2000 (first entry)
 XX
 DE Human cancer associated antigen precursor DNA, clone NY-REN-62.
 XX
 KW renal cancer; cancer associated antigen precursor; diagnosis;
 KW cytosolic; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 2..685
 FT /*tag= a
 FT /transl_except= (pos:38..43, aa:Gly)
 FT /transl_except= (pos:59..64, aa:Gly)
 FT /transl_except= (pos:281..286, aa:Ala)
 FT /transl_except= (pos:401..406, aa:Lys)
 FT /transl_except= (pos:620..625, aa:Lys)
 XX
 PN M0200020587-A2.
 XX
 PD 13-APR-2000.
 XX
 PF 04-OCT-1999; 99WO-US22873.
 XX
 PR 05-OCT-1998; 98US-0166300.
 PR 05-OCT-1998; 98US-0166350.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 XX
 PI Obata Y, Gout I, Tureci O, Sahin U, Pfeundschuh M, Scanlan MJ;
 PI Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
 XX
 DR WPI; 2000-303774/26.
 DR P-PSDB; AAY92346.
 XX
 PT Preventing, diagnosing and/or treating disorders associated with
 PT abnormal expression of human cancer associated antigens
 XX
 PS Claim 57; Page 84; 121pp; English.
 XX
 CC AAA09310-20 are novel genes isolated by SEREX screening from a renal
 CC cancer cell line 1973/10.4. The genes encode cancer associated antigen
 CC precursors. These gene products are useful in methods for preventing,
 CC diagnosing and/or treating disorders, especially cancer, associated with
 CC abnormal expression of human cancer associated antigens. The method
 CC comprises contacting a sample from a subject with an agent that
 CC specifically binds to the nucleic acid molecule or expression product
 CC (or fragment) complexed with a human leukocyte antigen (HLA) molecule
 CC and determining the interaction between the agent and the nucleic acid

CC molecule or the expression product as a determination of the disorder.
XX
SQ Sequence 686 BP; 299 A; 85 C; 171 G; 126 T; 5 other;

Query Match 1.0%; Score 40.4; DB 21; Length 686;
Best Local Similarity 50.3%; Pred. No. 0.53;

Matches 98; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

OY 1956 ctccgcaaggctctccagctgagcatgatattcccaattgtctgtgacctcatgagca 2015

DB 472 ctccgtagaattctactcagaagaagcaaaaagttgacctgataatgaaagaa 531

OY 2016 gaggctaccaggaagaaagcaaaacttgaaagtactgctgtaaaacagacat 2075

DB 532 actccaagcccaagcaagaactgcagactccaagcactcaaaaagcaagcagcaag 591

OY 2076 gatagatgacatcccaagatatgttgtaactgctagcttaaaaaccagatgagca 2135

DB 592 gtctgcttaaaatcagctcagtgatgaaagcnaatgaaagtgcagcagatgtgat 651

OY 2136 gcttatgactgagac 2150

DB 652 ggaatgaaaaaac 666

RESULT 9

ABL34172
ID ABL34172 standard; DNA; 15518 BP.

XX ABL34172;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 2145.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; anti-anemic; cytosine; cytosine; cytosine;

KW antiarteriosclerotic; anti-anemic; cytosine; cytosine; cytosine;

KW antiarteriosclerotic; anti-anemic; cytosine; cytosine; cytosine;

KW antiarteriosclerotic; anti-anemic; cytosine; cytosine; cytosine;

OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

XX for diagnosis and treatment of diseases associated with abnormal

XX cytosine methylation.

XX Claim 1; SEQ ID NO 2145; 32bp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

XX genes which are modified by the methylation of cytosines. The sequences

XX can be used in the diagnosis and treatment of immune system disorders,

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 15518 BP; 3047 A; 362 C; 4183 G; 7926 T; 0 other;

Query Match 1.0%; Score 40.2; DB 24; Length 15518;
Best Local Similarity 51.6%; Pred. No. 4.2;

Matches 116; Conservative 0; Mismatches 108; Indels 1; Gaps 1;

OY 60 tgaattgattgctcttgatataccagagctgctgttggcttggcttggcttggctt 119

DB 14825 tgggtgttggcttggcttggcttggcttggcttggcttggcttggcttggctt 14884

OY 120 ttgtgcttgatgcttgatgcttgatgcttgatgcttgatgcttgatgcttgatg 179

DB 14885 gaggtagtgcttgcttgatgcttgatgcttgatgcttgatgcttgatgcttgatg 14944

OY 180 ttgagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 239

DB 14945 ttgggttgatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 15003

OY 240 tatgagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 284

DB 15004 gtttcagtttttggcttggcttggcttggcttggcttggcttggcttggcttgg 15048

RESULT 10

ABL34624
ID ABL34624 standard; DNA; 15518 BP.

XX ABL34624;

DT 26-MAR-2002 (first entry)

XX Human metastasis associated gene SEQ ID NO: 177.

KW Metastasis associated gene; cytosine; gene therapy; cancer;

KW cytosine methylation; gene; ds.

OS Homo sapiens.

XX WO200177376-A2.

XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-EP03970.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-010922/01.

XX New nucleic acid derived from chemically treated metastasis genes,

XX useful for diagnosis of cancers by analysis of cytosine methylation,

XX also for treatment.

XX Claim 1; SEQ ID NO 177; 23bp + Sequence Listing; English.

XX The present invention provides a number of human metastasis associated

XX genes which are modified by cytosine methylation. The sequences can

XX be used in the diagnosis and treatment of cancer. The present sequence is

XX one of the genes of the invention.

XX Sequence 15518 BP; 3047 A; 362 C; 4183 G; 7926 T; 0 other;

	Query Match	1.0%	Score 40.2;	DB 24;	Length 15518;
	Best Local Similarity	51.6%;	Prod. No. 4.2;		
	Matches 116;	Conservative	0;	Mismatches 108;	Indels 1;
					Gaps 1;
Oy	60	tgatgatgtctcttcgtgatataccagagctggttttggtttgtggtttgtgt	119		
Db	14825	tgtgtgtttgtgtgtgtgtgtttttgttaagggttcgttgtagttgtttgttggtt	14804		
Oy	120	ttgtgcttggaattgtttgabtgcataatccggcggttaacagatacaactgcgtgattgata	179		
Db	14885	gagtcgggtgtgtgttcgtcatatgttcgaagtcctggagatttttttttttgcgttcgttag	14944		
Oy	180	ttgagtttgcctcgcgcgtgtctgtgtgttttaattccctccatgcgtgtgatacca	239		
Db	14945	ttgggttttagtgt-agtgtgtgtgtttgatttaattttttgtcagtagagatag	15003		
Oy	240	tatgagagattgttcgacgtagatcacgagaggagctcggtgtttgttg	284		
Db	15004	gttttagtttttttccttaaggttaagttcgaatggaaggaattttaaagg	15048		

RESULT	11
ABL32775/c	
ID	ABL32775 standard; DNA; 15122 BP.

DT	26-MAR-2002	(first entry)
XX		
DE	Human immune system associated gene SEQ ID NO: 748.	

Human, immune system disease; cytosine methylation; antiasclerotic; antiartherosclerotic; antianemia; cytosolic; neurologic; neuroprotective; anti-HIV; anticonvulsant; ophthalmologic; antineumatic; antiarthritic; antidiabetic; antipsoriatic; antineoplastic; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; Rheumatoid arthritis; psoriasis; bowel disease; gene; ds.

OS Homo sapiens.

PN WQ200200928-A2.

PD 03-JAN-2002.

02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -

PS Claim 1; SEQ ID NO 748; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/allergic bowel diseases. The present sequence is a gene of the invention.

Sequence 15122 BP; 3782 A; 385 C; 3759 G; 7196 T; 0 other; 50

Query Match	1.0%;	Score 39.8;	DB 24;	Length 1512;
Best Local Similarity	50.8%;	Pred. No. 5.4;		
Matches 95;	Conservative 0;	Mismatches 92;	Indels 0;	Gaps 0;

RESULT	12
AAH54703	
ID	AAH54703 standard; DNA; 3257 BP.

DT	03-SEP-2001 (first entry)
XX	
DE	S. epidermidis genomic polynucleotide sequence SEQ ID NO:4067.

staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination; endocarditis; ds.

Staphylococcus epidermidis.

PN W0200134809-A2.

17-MAY-2001
PD

09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PA (GLAX) GLAXO GROUP LTD.

Kimmerly WJ;

DR WPI; 2001-316495/33.

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1752-1753; 2188pp; English.

Claim 8; Page 1752-1753; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (1) encoding polypeptides (11), given in AGA81454 to AA683120, from *Staphylococcus epidermidis*. CC (1) and (11) can have antibacterial activity and therefore can be used CC in vaccination. The nucleic acids (1) may be used to produce the CC S. epidermidis polypeptides (11) via the production of vectors CC containing them which are used to produce hosts cells which express the CC polypeptides. The polypeptides (11) (and/or nucleic acids) may then be CC used to vaccinate subjects and to raise antibodies against the bacteria. CC The polypeptides may also be used to assay for other inhibitors of their CC activity and therefore identify compounds that may be used for the CC treatment of S. epidermidis infections, e.g. endocarditis. AAH5371 to CC AAH5090 represent specifically claimed S. epidermidis genomic DNA CC polynucleotide sequences from the present invention. AAH5091 to CC AAH5098 represent oligonucleotide sequences and primers which are used CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454, so even
 CC though sequences are given in the disclosure for SEQ ID NO:4455 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

SQ Sequence 3257 BP; 1140 A; 462 C; 586 G; 1069 T; 0 other;

Query Match 1.0%; Score 39.6; DB 22; Length 3257;
 Best Local Similarity 50.5%; Pred. No. 2.4;
 Matches 96; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

OY 724 aaggaagaatctgatacactcttccaagaagtgtaagaaggaatgactca 793
 DB 22 aagcaaaaaacagagagcgctgtgatacgttcaattatagaagaagcatatgaacatca 81
 OY 794 aatgcaatgcgcctctctgcaagaatgagctgctgagccatactgattcaccatga 853
 DB 82 tgtatgatactctctctctgaaatgtaagtaagtagtacttactgacaacatca 141
 OY 854 aagatttcgaagggccagcccaaatatgattgctgagcaaatgctctgaggaaca 913
 DB 142 atgaagaatgagcctgaattatgattgattgaaagaatctaattaaagaacaaga 201
 OY 914 cttctgttga 923
 DB 202 catcgctta 211

RESULT 13
 AAS92080/C
 ID AAS92080 standard; CDNA; 843 BP.

AC AAS92080;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #27884.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG27893.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 1; SEQ ID NO 27884; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 843 BP; 182 A; 212 C; 190 G; 259 T; 0 other;

Query Match 1.0%; Score 39.4; DB 23; Length 843;
 Best Local Similarity 54.5%; Pred. No. 1.2;
 Matches 79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 1207 tctgaagaagttgtctcaagaagctccaatctcaagaggaagcagataagaattg 1266
 DB 252 tctgatatcagacacgtaggaaaaacagaaagaaagaaagaaagaaagaaagaaag 193
 OY 1267 atgaagaagcagcagcacagcaagaacgcactgcccagctgatttccaagtgcaag 1326
 DB 192 AAGAAGACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 133
 OY 1327 cttgtcggagaagaccacaaaagg 1351
 DB 132 GGTGTGATCTTCAGGCCAAGCAATG 108

RESULT 14

AA238223
 ID AA238223 standard; DNA; 2167 BP.

AC AA238223;

DT 09-FEB-2000 (first entry)

DE Human S184L mutant protein kinase MKK4 gene fragment E.

XX MKK4; mitogen activated protein kinase; MAPK; MAPK pathway; mutation;

KW somatic; signal transduction; apoptosis; stress; cytokine; induction;

KW phosphorylation; Jun kinase; JNK; p38; tumour; suppressor; mutant;

KW loss of heterozygosity; LOH; cancer; detection; diagnosis; prognosis;

KW breast cancer; pancreatic cancer; colorectal cancer; testicular cancer;

KW drug screening; gene therapy; protein replacement therapy; mimetic; ds.

XX Homo sapiens.

OS Synthetic.

PI Key Location/Qualifiers

PH Key 1..247

FT intron /*tag= a

FT exon 248..367

FT /*tag= b

FT /*note= "MKK4 mutant exon E"

FT intron 368..2167

FT /*tag= c

PN US989885-A.

PD 23-NOV-1999.

PF 13-JUN-1997; 97US-0874186.

PR 10-JAN-1997; 97US-0782482.

Tue, Sep 17 07:37:04 2002

us-09-828-068-1.rng

Page 11

Db 2282 atagctggagcccttgctcatctctgaggagccctctacagcatttaagttatga 2333

Search completed: September 16, 2002, 20:19:54
Job time: 12101 sec

THIS PAGE BLANK (USPTO)

MOLECULE TYPE: CDNA
US-09-098-487-2

Query Match 0.9%; Score 36.4; DB 2; Length 2277;
Best Local Similarity 29.7%; Pred. No. 0.81;
Matches 116; Conservative 66; Mismatches 206; Indels 3; Gaps 2;

QY 908 acaacactctgttgatgttgaggcttaacgtgaagttcccccagattacatgagcagatg 967
DB 391 AARACNGTWTYGCNCARTTYGNGCNGTNTYNGARGTMAAATATCCNMGNARCCNGAY 450
QY 968 aagtaaatgtgacagatcaacatccatccacacccaactctcgaatgagtcctcaaaa 1027
DB 451 GGNAAATGTGNGGNTTGTGNTTGTNCARTTYAARAAATYTYTNGARCCNGNARCGN 510
QY 1028 gaattgaagatgaaaaatggaactgaagagac-tcttgctgctgagcagtgcaattg 1086
DB 511 YTNAAAGNATGAAYATGAARATHAARGMGNACNGTNGCNGTNGAYTGGCGNGTN 570
QY 1087 accaagatcctaaccaatgtctgaaaggaacgtgacagtgctgagcagtgcaat 1146
DB 571 GCNAAARCAVYARPYARCAVYACNCARMSNGTWSNGCATHGNGARARARARMSNCAY 630
QY 1147 ttgaccaaagatccgaacacagtgctgagcagaatgtgaagcagatctgcaatgagcca 1206
DB 631 GARMSNARCAVYCARGARMSNGTNAARARARAGMNGARGARARGARAYATGARGAR 690
QY 1207 tctgaagaagtgttctccaaagaagatccaatctaaagagagaagcagtaagaatgtg 1266
DB 691 GARARARAYAGV--ATGATGATGATGATGATGATGATGATGATGATGATGATGATG 748
QY 1267 atgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1297
DB 749 ARGAYGARAGARGARARARATHGARMSNMA 779

RESULT 11
US-08-510-133A-32
Sequence 32, Application US/08510133A
Patent No. 6221839

GENERAL INFORMATION:

APPLICANT: Alltalo, Karl
TITLE OF INVENTION: Receptor Ligand
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,133A
FILING DATE: 01-Aug-1995
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 28113/32863
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 37..1089
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-08-510-133A-32

Query Match 0.9%; Score 36.2; DB 4; Length 1140;
Best Local Similarity 53.1%; Pred. No. 0.56;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1741 aatgataagacagatgtctgcaagatgtatcagaatctccacacagaggtgctca 1800
DB 827 AATTTGATGAAACACATGCGAGTGTATGTAAAGAACCTGCCAGAAATCAACCCC 886
QY 1801 tcaaggggaaacagcgggttgagtaagggaacacattcagctgtatgacaaa 1860
DB 887 TAAATCTGGAATATGTCCTGTGATGTACAGAAATCCACAGAAATGCTGTAAAG 946
QY 1861 tatgtgtgaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1885
DB 947 GAAAGAGTTCACACACCAACATG 971

RESULT 12
US-08-585-895-32
Sequence 32, Application US/08585895
Patent No. 6245530

GENERAL INFORMATION:

APPLICANT: Alltalo, Karl
TITLE OF INVENTION: Receptor Ligand
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/585,895
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 28113/33072
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 37..1089
NAME/KEY: mat_peptide

RESULT 15
US-09-042-105-3
Sequence 3, Application US/09042105
Patent No. 6040157
GENERAL INFORMATION:
APPLICANT: HU, JING-SHAN
APPLICANT: ROSEN, CRAIG A.
APPLICANT: CAO, LIANG
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR 2
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX
STREET: 1100 NEW YORK AVENUE
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042.105
FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/207,550
FILING DATE: 8-MAR-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,968
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: TO BE ASSIGNED
FILING DATE: 24-DEC-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: ERIC K. STEFFE
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1000003/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1526 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 71..142
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 143..1120
FEATURE:
NAME/KEY: CDS
LOCATION: 71..1120
US-09-042-105-3

Query Match 0.98; Score 36.2; DB 3; Length 1526;
Best Local Similarity 53.18; Pred. No. 0.7; Mismatches 68; Indels 0; Gaps 0;
Matches 77; Conservative 0;

QY 1741 aatagcataagaagatctgtcagcatgtatcagaatctccacacagagtgctca 1800
DB 861 AATTGATGATAAACAATGCCAGTGTATGTAAAGAAGAACTGCCCAAGAAATCAACCCC 920
QY 1801 tcaaaagggaacacagcggttgtagtaagggaacacacattcagctgctagtaccaaa 1860

DB 921 TAAATCCTGGAATAATGTGCTGTGAATGTACACAAAGTCCACAGAAATGCTGTAAAG 980
QY 1861 taigtggtgaagccaccagaaatg 1885
DB 981 GAAGAAGTTCACACCAACAAATG 1005

Search completed: September 16, 2002, 20:09:49
Job time: 11506 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 15:26:02 ; Search time 3935.68 seconds
(without alignments)
13360.877 Million cell updates/sec

Title: US-09-828-068-1
Perfect score: 3896
Sequence: 1 cgcgcgtcgtcgcgcgaacacg.....gtcaacacgcgcgcgaatttacc 3896

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1336207 seqs, 674847542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79.8	2.0	321	9	AI691520
2	61.8	1.6	895	12	CNS0071A
3	57.8	1.5	1101	12	CNS0039G
4	56.2	1.4	922	12	CNS0073W
5	54	1.4	1101	12	CNS016LW
6	53.8	1.4	884	12	CNS006U0
7	52.8	1.4	939	12	CNS00CNG
8	51.8	1.3	1101	12	CNS00DG1
9	50.4	1.3	997	12	CNS005TE
10	48.6	1.2	909	12	CNS00JTL
11	47.4	1.2	985	12	CNS017TG
12	46.4	1.2	1101	12	CNS00LO0
13	46.2	1.2	1101	12	CNS017WI
14	46	1.2	706	9	AM684844
15	45.6	1.2	748	10	BI176637
16	45.6	1.2	1101	12	CNS00PXE
17	45.6	1.2	1101	12	CNS00LT2

c 18	45.2	1.2	490	12	A0584620
c 19	45.2	1.2	918	12	AG126399
c 20	45	1.2	937	12	CNS006ST
c 21	44.8	1.1	902	12	CNS006OP
c 22	44.6	1.1	450	9	AU060996
c 23	44.4	1.1	413	9	AI641313
c 24	44.2	1.1	1134	12	AG130678
c 25	44	1.1	860	12	CNS018FL
c 26	44	1.1	861	12	CNS0075A
c 27	44	1.1	941	12	BH156851
c 28	43.8	1.1	826	10	BF242104
c 29	43.6	1.1	1101	12	CNS017ET
c 30	43.4	1.1	754	10	BF845969
c 31	43.4	1.1	1101	12	CNS00GPB
c 32	43.2	1.1	1043	12	AG135568
c 33	42.6	1.1	517	12	AG127302
c 34	42.2	1.1	233	12	A2267626
c 35	42.2	1.1	1781	9	AW727582
c 36	42	1.1	530	10	C93179
c 37	42	1.1	624	12	A2031624
c 38	42	1.1	684	9	AW187871
c 39	42	1.1	943	12	CNS05T26
c 40	41.8	1.1	513	12	CNS02CMF
c 41	41.8	1.1	640	10	BE427036
c 42	41.8	1.1	650	10	BE427548
c 43	41.8	1.1	905	12	A2550256
c 44	41.8	1.1	1058	10	BE427089
c 45	41.6	1.1	819	12	A0744726

ALIGNMENTS

RESULT 1
AI691520/c 321 bp mRNA linear EST 02-FEB-2000
LOCUS 606020C11.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION AI691520
VERSION AI691520.1 GI:4966664
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 321)
REFERENCE
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606020 row: C column: 11.
Location/Qualifiers
1..321
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/note="Organ: Immature ear; Vector: pRK-CMV; Site 1: EcoRI
; Site 2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"
BASE COUNT 78 a 91 c 59 g 91 t 2 others

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT									
Db	1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1603	Db	491	AAAAA
1031	1124	971	1184	911	1244	851	1303	791	1363	731	1423	671	1483	611	1543	551	Qy	1			

filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES
    source
        1..922
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_1lb="RPCT-98"
            /clone="BACRIAD09"
            /note="end : TET3"
BASE COUNT      223 a      95 c      109 g      221 t      274 others
ORIGIN

Query Match      1.4% Score 56.2; DB 12; Length 922;
Best Local Similarity 21.6%; Pred. No. 0.0031;
Matches 66; Conservative 120; Mismatches 119; Indels 1; Gaps 1;

OY 35 tgcctacgcgtgtgcaagcgcgtatgattgctctcgtgatatccagagctcg 94
    :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 617 KCTKKTTTKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 676
    KCTKKTTTKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK

OY 95 tgttttgagtc-tgtggttttggttttggttgagatgtgagtcgaatccgcgc 153
    :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 677 KTTGKKKKGGKTKTKTKTKTKTKGKTKTKTKTKTKTKTKTKTKTKTKTK 736
    KTTGKKKKGGKTKTKTKTKTKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTK

OY 154 gttaacaagactactcgtatgatcatatgattgtgcgcgcgcgtgtgctgtgt 213
    :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 737 KTTTNTTKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 796
    KTTTNTTKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK

OY 214 gattctctcctgcgtgtgtgattgatatgagatgtgtgcagtagatcaggaggagct 273
    :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 797 KCTGTKKKTKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 856
    KCTGTKKKTKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK

OY 274 cgtgttgttggagcaagactgtatgctctgtcgtgtgtggaactggtgtctagcgcagt 333
    :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 857 KTTGTGTTTTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 916
    KTTGTGTTTTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

OY 334 ttggag 339
    :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 917 GKKKKG 922
    :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 5
CNS016LM 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BAC1616 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106910
VERSION AL106910.1 GI:5624430
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
GENOSCOPE.
Direct Submisson
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES
    source
        1..1101
            /organism="Drosophila melanogaster"
            Location/Qualifiers
```

The BACP is constructing a reference melanosaster genome using these BACs. For further information please see <http://www.fruitfly.org> The Drosophila melanosaster BAC library was prepared by Kazuhiro Oseogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

ACCESSION	fly), genomic survey sequence.					
VERSION	AL108478					
KEYWORDS	GSS.					
SOURCE	fruit fly.					
ORGANISM	<i>Drosophila melanogaster</i>					
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> .					
AUTHORS	1 (bases 1 to 985)					
TITLE	Genoscope.					
JOURNAL	Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the European <i>Drosophila</i> Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This <i>Drosophila melanogaster</i> BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobac11.					
FEATURES	location/Qualifiers					
SOURCE	1..985					
	/organism="Drosophila melanogaster"					
	/plasmid="pBelobac11"					
	/db_xref="taxon:7227"					
	/clone_lib="DrosBAC"					
	/clone="BACN37G05"					
	/note="end : SP6"					
BASE COUNT	139 a 188 c 248 g 115 t 295 others					
ORIGIN						
Query Match	1.28; Score 47.4; DB 12; Length 985;					
Best Local Similarity	36.18; Pred. No. 0.59;					
Matches	57; Conservative 42; Mismatches 59; Indels 0; Gaps 0;					
Oy	32 ggtttcttaagcgttgcaagcgctagctgattgatttccttgcgtgatatccagaagc 91					
Db	811 GTTGTCTWTTMKTKTKKGGGGGGGGGGGCGTWTTTTTTKTKTTTKTKTGTTTKTKTGKT 870					
Oy	92 tcgttgtttggttggttggttggttggttggttcgttcgttcgattgattgattgcctaactgcg 151					
Db	871 TTGCTGCTTKTTTGATKKKKKKTKGKKTKGKKTKGKKKGAGKGGTGTTGTTGKTKTGTCATG 930					
Oy	152 gcggttaacaagatcacctgctgattgattgaatatgaattgtg 189					
Db	931 TGKKKGKGGKGYTKTKKKGCTGTTTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 968					
RESULT 12						
CNS00LO/c	1101 bp DNA linear GSS 03-JUN-1999					
LOCUS						
DEFINITION	<i>Drosophila melanogaster</i> genome survey sequence TET3 end of BAC: BACR3J023 of RPCI-98 library from <i>Drosophila melanogaster</i> (fruit fly), genomic survey sequence.					
ACCESSION	AL068607					
VERSION	AL068607.1 GI:4958689					
KEYWORDS	GSS.					
SOURCE	fruit fly.					
ORGANISM	<i>Drosophila melanogaster</i>					
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> .					
REFERENCE	1 (bases 1 to 1101)					
AUTHORS	Genoscope.					
TITLE	Direct Submission					
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a					
COMMENT						

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Oseegawa and Aaron Kammasser in Pieter de Jong's Laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES		location/Qualifiers
source	1..1101	/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone_1lb="RPCT-98"
		/clone="BACR32D23"
		/note="end : Ter3"
BASE COUNT	155 a 166 c 7 g 284 t 489 others	
ORIGIN		
Query Match	1.2%; Score 46.4; DB 12; Length 1101;	
Best Local Similarity	21.0% ; Pred.No.1.1;	
Matches 65; Conservative 112; Mismatches 132; Indels 0; Gaps 0;		
Oy	75	tgtgatataccagggcctcgatgatttggttgcgttcgttttgttggttcgttcgttatgattgt 134 :: : : : : : : : : : : : : :
Dd	810	TKKKGRKKKKKTGGGAKRTTMMTKTTTGKGtGTAMKTGTGKKKKKKGGKKGMWTGt 751
Oy	135	tgaatgctaactcccggcgcatcaagaacacctgtagatatgataltaagtltgcccg 194 : : : : 750 KMDTMTTKTKGKTTSGCGKKKTGYKTYKAAMAANKAKTKTTTTKKKGATADTKKK 691
Oy	195	gctgtgcgtgcgtgtlgttgatctctctccgcgctgvtgtagatagaraggatgttcgc 254 :: : Db 690 KKKDAGTKTTGGKKKKTKGAAMWGTDTRDKKKRAMAAMNDVTGKTFRAKKRADPKDTG 631
Oy	255	agtaaacaaggsgagaccgcgtgtgttvggscagaaactgatcctgtccgtcggtfgaac 314 : : : : Db 630 KGAAATATAKAKKGDADADGTDITKKAKTGTGTKADDTKGGAARGGKRKXTKDKGA 571
Oy	315	tgtgtcgtfagagcacagitttgagctacagacagcccccttcagatgcagcccgta 374 : ::~ :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : Db 570 KDDKDATAGAkkkgkTAkKgDAkKkdGKaDGARtkRKkKkAcgTKKdTdkKdgmdK 511
Oy	375	agctgtgctg 383 :: :
Db	510	KGAtRGtGtK 502
RESULT 13	CNSO17M/c	1101 bp DNA linear GSS 26-JUL-1999
LOCUS	CNSO17MI	
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BAONJ3L10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL108588	
VERSION	ALI08588.1 GI:5628892	
KEYWORDS	GSS,	
SOURCE	Fruit fly.	
ORGANISM	Drosophilinae melano-gaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophilla.	
REFERENCE	Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :	
AUTHORS	BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	
TITLE	JOURNAL	

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2002, 22:23:21 ; Search time 88.88 Seconds

(Without alignments)
2057.332 Million cell updates/sec

Title: US-09-828-068-2

Perfect score: 5526

Sequence: 1 MEIVAVDQEGARVGTNCML.....NKNPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5526	100.0	1385	10	Q9LW14
2	5516	99.8	1057	10	Q947D1
3	393	7.1	1096	10	Q9LYD9
4	221	4.0	1983	4	Q9Y4D6
5	216	3.9	2081	10	Q9LH98
6	215.5	3.9	3507	5	Q23587
7	215	3.9	1690	5	Q44929
8	209.5	3.8	852	10	Q9S255
9	208.5	3.8	1295	5	Q22257
10	207.5	3.8	1520	12	Q9QJ16
11	206	3.7	3484	5	P91257
12	205.5	3.7	2083	5	Q9N435
13	204.5	3.7	5327	5	Q76891
14	204	3.7	2768	5	Q9VC00
15	204	3.7	2951	5	Q9W320
16	203	3.7	1047	5	Q9GKR7

17	202.5	3.7	2271	16	Q990Y4	Q99Y4 staphylococ
18	202	3.7	2062	4	Q9H231	Q9H231 homo sapien
19	201.5	3.6	665	5	Q96229	Q96229 plasmodium
20	201.5	3.6	1784	5	Q9VE02	Q9VE02 dirosophila
21	200.5	3.6	3254	5	Q9BK45	Q9BK45 plasmodium
22	200	3.6	2647	5	Q9U4X0	Q9U4X0 plasmodium
23	200	3.6	3111	5	Q9VH10	Q9VH10 dirosophila
24	199.5	3.6	1302	2	Q49547	Q49547 mycoplasma
25	199.5	3.6	17352	5	Q951W2	Q951W2 procamburus
26	199	3.6	1230	5	Q20626	Q20626 caenorhabd1
27	198.5	3.6	1043	10	Q82345	Q82345 aradidops1s
28	198.5	3.6	1690	5	Q9VJES	Q9VJES dirosophila
29	198	3.6	615	6	Q77733	Q77733 saginus oe
30	197.5	3.6	1852	3	Q9C2H4	Q9C2H4 neurospora
31	196.5	3.6	963	11	Q99M44	Q99M44 glaucomyx v
32	196.5	3.6	1579	13	Q90X36	Q90X36 xenopus lae
33	196	3.5	1513	12	Q9WS28	Q9WS28 human herpe
34	196	3.5	2910	10	Q9FND5	Q9FND5 aradidops1s
35	195	3.5	667	4	Q96D08	Q96D08 homo sapien
36	195	3.5	1616	4	Q96PH3	Q96PH3 homo sapien
37	195	3.5	1730	5	Q9VC59	Q9VC59 dirosophila
38	193.5	3.5	1803	5	Q9VU16	Q9VU16 dirosophila
39	192	3.5	3201	5	Q9W0U2	Q9W0U2 dirosophila
40	192	3.5	6632	5	Q17362	Q17362 caenorhabd1
41	192	3.5	6642	5	Q01761	Q01761 caenorhabd1
42	191.5	3.5	990	13	Q91803	Q91803 xenopus lae
43	191.5	3.5	2158	4	Q9Y6Y3	Q9Y6Y3 homo sapien
44	191.5	3.5	2265	4	Q9Y6Y4	Q9Y6Y4 homo sapien
45	191.5	3.5	3394	4	Q9Y6V0	Q9Y6V0 homo sapien

ALIGNMENTS

RESULT 1

ID Q9LW14 PRELIMINARY: PRT: 1385 AA.

AC Q9LW14: 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DI 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE HYPOTHETICAL PROTEIN.

OS Oryza sativa (Rice).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: OC Euphorbiaceae: Oryzae; Oryza.

OX NCBI_Taxid=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nippondare(GAS) genomic DNA, chromosome 1, PAC clone: P0485D09.";

RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AP001859; BAA94774.1; -

KW Hypothetical protein.

SO SEQUENCE 1385 AA; 151678 MW; 16E784264EEFC7B5 CRC64;

Query Match 100.0%; Score 5526; DB 10; Length 1385;

Best local Similarity 100.0%; Pred. No. 0;

Matches 1057; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEIVAVDQEGARVGTNCMLARGGTGAVAVYLELTTPRODAAAEKGVDEPAHOCEHFS 60

DB 329 MEIVAVDQEGARVGTNCMLARGGTGAVAVYLELTTPRODAAAEKGVDEPAHOCEHFS 368

QY 61 IRGVAVLLQKKDPKFCFSIRIFHDQKKCEHKAASSPFAKRRMDCKCLDKLTSDN 120

DB 389 IRGVAVLLQKKDPKFCFSIRIFHDQKKCEHKAASSPFAKRRMDCKCLDKLTSDN 448

QY 121 GAAPTLPAKQNGTSDGCSITFEVRSFTVPASVSGKVSPTOSGCKMADRSTLPKSVQE 180

Db 449 GTAARLPAKONGTSDGCSITFFVSTFVPASVSGOKVSPSTOSSQGNADRLTPKSVOE 508
 181 GNSCKNAPSGKNGAANANDSPMKDLQGPQAVYDVAANYSEDTSDVAGALPEVQITW 240
 509 GNSCKNAPSGKNGAANANDSPMKDLQGPQAVYDVAANYSEDTSDVAGALPEVQITW 568
 241 HIEVNGADQPPSTPKLSEVYLAKRNEDENGKTEETLVAEQNLTKDPNPSGKERDQVAEQ 300
 569 HIEVNGADQPPSTPKLSEVYLAKRNEDENGKTEETLVAEQNLTKDPNPSGKERDQVAEQ 628
 301 CNLTKDPKPVSGQCEQICNEPCEEVYLLKSSSKKRTDKMLKKQOHSKKRTAQADAVSD 360
 629 CNLTKDPKPVSGQCEQICNEPCEEVYLLKSSSKKRTDKMLKKQOHSKKRTAQADAVSD 688
 361 AKLCRRKPKVRLSEITINANOVEDSRSDVEYHRENAADPCEDDRSTIPVMEVSMIDPV 420
 689 AKLCRRKPKVRLSEITINANOVEDSRSDVEYHRENAADPCEDDRSTIPVMEVSMIDPV 748
 421 NHVGEDGLKSSKNKTKRKYSDDVDDGSSLMMNLNGKKKRTGSHHTVAHPAGNLSNKKV 480
 749 NHVGEDGLKSSKNKTKRKYSDDVDDGSSLMMNLNGKKKRTGSHHTVAHPAGNLSNKKV 808
 481 TPTASTQHDDENDENGIDTMMHKTQVCOHYSEISTORCSKGTAGLSKGTSAASTK 540
 809 TPTASTQHDDENDENGIDTMMHKTQVCOHYSEISTORCSKGTAGLSKGTSAASTK 868
 541 YGGESTRNGQNIHVLASADQCOMETENSVLSHSAKVSPEAHDIOIMSDLHQSLEPKKKK 600
 869 YGGESTRNGQNIHVLASADQCOMETENSVLSHSAKVSPEAHDIOIMSDLHQSLEPKKKK 928
 601 QKLEVTREKQMTIDIPMDIYELLAKNOHERQMLTETDCSDINRIQSTTADDDCVIYA 660
 929 QKLEVTREKQMTIDIPMDIYELLAKNOHERQMLTETDCSDINRIQSTTADDDCVIYA 988
 661 KGSQYASSVFPTNSQOKSLASQSTOKELQGHIALTQESPHQNFQSTOEQOHTLRME 720
 989 KGSQYASSVFPTNSQOKSLASQSTOKELQGHIALTQESPHQNFQSTOEQOHTLRME 1048
 721 MYTIAASSPLFESHDDQYIAEAPTEHMGKDAKLTWQFKATTNSPAATCGAOFRG 780
 1049 MYTIAASSPLFESHDDQYIAEAPTEHMGKDAKLTWQFKATTNSPAATCGAOFRG 1108
 781 QAVDLSTHVMGSSSNYASROPVIALDRYARAVNOVYARNFPTIATMEASKLCDRN 840
 1109 QAVDLSTHVMGSSSNYASROPVIALDRYARAVNOVYARNFPTIATMEASKLCDRN 1168
 841 AGQVLLPKESMPATHILRMMDSTLASFPNYGTSRNQMESQLHNSQYAHNOYKGSST 900
 1169 AGQVLLPKESMPATHILRMMDSTLASFPNYGTSRNQMESQLHNSQYAHNOYKGSST 1228
 901 SYGSNLNGKIPLTFFEDLSRHQJLHRLRPHRPGVGLSLQKELANSENKGTQSGYK 960
 1229 SYGSNLNGKIPLTFFEDLSRHQJLHRLRPHRPGVGLSLQKELANSENKGTQSGYK 1288
 961 LGVSTGITSNOMRKKEHFEALNSGMFSAKMNAQLQGSVSSADFLSARNSIAQSWTRGK 1020
 1289 LGVSTGITSNOMRKKEHFEALNSGMFSAKMNAQLQGSVSSADFLSARNSIAQSWTRGK 1348
 1021 KAVHPLDRFVRQDICTTKNKPADFTTISNDNEMDR 1057
 1349 KAVHPLDRFVRQDICTTKNKPADFTTISNDNEMDR 1385
 RESULT 2
 ID 0947D1 PRELIMINARY; PRT; 1057 AA.
 AC 0947D1;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE EMBRYONIC FLOWER 1-LIKE PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriaraloideae; Oryzaeae; Oryza.
 OX NCBI_TaxID=4330;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21380417; PubMed=11487698;
 RA Aubert D., Chen L., Moon Y.H., Martin D., Castle L.A., Yang C.H.,
 RA Sung Z.R.;
 RT "Emf1, a novel protein involved in the control of shoot architecture
 and flowering in arabidopsis."
 RL Plant Cell 13:1865-1875(2001).
 DR EMBL: AF326768; AAK98529.1;
 SO SEQUENCE 1057 AA; 116447 MW; 2A44FA15C93C3DA0 CRC64;

Query Match 99.8%; Score 5516; DB 10; Length 1057;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1055; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 MEIVADQEGARVYGTNCMLARGTGAVAPVLELTATPRQDAAEAGVDEPAQOCEHRS 60
 1 MEIVADQEGARVYGTNCMLARGTGAVAPVLELTATPRQDAAEAGVDEPAQOCEHRS 60
 61 IRGVALLQKKDKPFCSLRIFFHDKKDEHKAASSPFSYAKFRWDCKLDTKTSN 120
 61 IRGVALLQKKDKPFCSLRIFFHDKKDEHKAASSPFSYAKFRWDCKLDTKTSN 120
 121 GTAARLPAKONGTSDGCSITFFVSTFVPASVSGOKVSPSTOSSQGNADRLTPKSVOE 180
 121 GTAARLPAKONGTSDGCSITFFVSTFVPASVSGOKVSPSTOSSQGNADRLTPKSVOE 180
 121 GTAARLPAKONGTSDGCSITFFVSTFVPASVSGOKVSPSTOSSQGNADRLTPKSVOE 180
 181 GNSCKNAPSGKNGAANANDSPMKDLQGPQAVYDVAANYSEDTSDVAGALPEVQITW 240
 181 GNSCKNAPSGKNGAANANDSPMKDLQGPQAVYDVAANYSEDTSDVAGALPEVQITW 240
 241 HIEVNGADQPPSTPKLSEVYLAKRNEDENGKTEETLVAEQNLTKDPNPSGKERDQVAEQ 300
 241 HIEVNGADQPPSTPKLSEVYLAKRNEDENGKTEETLVAEQNLTKDPNPSGKERDQVAEQ 300
 301 CNLTKDPKPVSGQCEQICNEPCEEVYLLKSSSKKRTDKMLKKQOHSKKRTAQADAVSD 360
 301 CNLTKDPKPVSGQCEQICNEPCEEVYLLKSSSKKRTDKMLKKQOHSKKRTAQADAVSD 360
 301 CNLTKDPKPVSGQCEQICNEPCEEVYLLKSSSKKRTDKMLKKQOHSKKRTAQADAVSD 360
 361 AKLCRRKPKVRLSEITINANOVEDSRSDVEYHRENAADPCEDDRSTIPVMEVSMIDPV 420
 361 AKLCRRKPKVRLSEITINANOVEDSRSDVEYHRENAADPCEDDRSTIPVMEVSMIDPV 420
 421 NHVGEDGLKSSKNKTKRKYSDDVDDGSSLMMNLNGKKKRTGSHHTVAHPAGNLSNKKV 480
 421 NHVGEDGLKSSKNKTKRKYSDDVDDGSSLMMNLNGKKKRTGSHHTVAHPAGNLSNKKV 480
 481 TPTASTQHDDENDENGIDTMMHKTQVCOHYSEISTORCSKGTAGLSKGTSAASTK 540
 481 TPTASTQHDDENDENGIDTMMHKTQVCOHYSEISTORCSKGTAGLSKGTSAASTK 540
 541 YGGESTRNGQNIHVLASADQCOMETENSVLSHSAKVSPEAHDIOIMSDLHQSLEPKKKK 600
 541 YGGESTRNGQNIHVLASADQCOMETENSVLSHSAKVSPEAHDIOIMSDLHQSLEPKKKK 600
 601 QKLEVTREKQMTIDIPMDIYELLAKNOHERQMLTETDCSDINRIQSTTADDDCVIYA 660
 601 QKLEVTREKQMTIDIPMDIYELLAKNOHERQMLTETDCSDINRIQSTTADDDCVIYA 660
 661 KGSQYASSVFPTNSQOKSLASQSTOKELQGHIALTQESPHQNFQSTOEQOHTLRME 720
 661 KGSQYASSVFPTNSQOKSLASQSTOKELQGHIALTQESPHQNFQSTOEQOHTLRME 720
 721 MYTIAASSPLFESHDDQYIAEAPTEHMGKDAKLTWQFKATTNSPAATCGAOFRG 780
 721 MYTIAASSPLFESHDDQYIAEAPTEHMGKDAKLTWQFKATTNSPAATCGAOFRG 780
 781 QAVDLSTHVMGSSSNYASROPVIALDRYARAVNOVYARNFPTIATMEASKLCDRN 840
 781 QAVDLSTHVMGSSSNYASROPVIALDRYARAVNOVYARNFPTIATMEASKLCDRN 840

```

Db 781 QAVDLSTHVMGSSSVYASROPVIAPLDRYAERAVNOVHARNPSTIATMEASKLCDRRN 840
QY 841 AGGVVLYPKESMPATILLRMDPSTLASPPNYCTSSRNQMESQJHNSQVHHNRYKSTST 900
Db 841 AGGVVLYPKESMPATILLRMDPSTLASPPNYCTSSRNQMESQJHNSQVHHNRYKSTST 900
QY 901 SVGSNNGKIPLETFEDLSRQHLDRPLRPHRVGLGSLQKEIAMSSENGTOSGYK 960
Db 901 SVGSNNGKIPLETFEDLSRQHLDRPLRPHRVGLGSLQKEIAMSSENGTOSGYK 960
QY 961 LGVSTGITHQNRKKEFEALNSGMFSAKMNAIQLSVSSSADFLSARNSIAOSWTRGKG 1020
Db 961 LGVSTGITHQNRKKEFEALNSGMFSAKMNAIQLSVSSSADFLSARNSIAOSWTRGKG 1020
QY 1021 KMVHPLDRFVRODICTNKNPADFTTISNDNEVMDYR 1057
Db 1021 KMVHPLDRFVRODICTNKNPADFTTISNDNEVMDYR 1057

RESULT 3
ID 09LYD9 PRELIMINARY; PRT: 1096 AA.
AC 09LYD9;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOHETICAL.121.7 KDA PROTEIN (EMBRYONIC FLOWER 1).
GN F15N18_120 OR EMF1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21380417; PubMed=11487698;
RA Aubert D., Chen L., Moon Y.H., Martin D., Castle L.A., Yang C.H.,
RA Sung Z.R.;
RT "Emf1, a novel protein involved in the control of shoot architecture
RT and flowering in arabidopsis."
RL Plant Cell 13:1865-1875(2001).
DR EMBL: AL163815; CAB87713.1;
DR EMBL: AF319968; AAK98528.1;
KW Hypothetical protein.
SQ SEQUENCE 1096 AA; 121670 MW; 190308C7A96A0F61 CRC64;

```

Query Match 7.1%; Score 393; DB 10; Length 1096;
 Best Local Similarity 21.0%; Pred. No. 6,4e-16;
 Matches 250; Conservative 189; Mismatches 403; Indels 348; Gaps 58;

```

Db 181 NNRSKKELNKPMSMDISSMKRQONVDAVTTFGSSEIAGVEDPPEPKATKHHKIGRLMEC 240
QY 267 ENKTEETTLVAEOCNLTKDPNPSGKER---DOVAEOCNLTKDPKPVSGKCEQICNEPC 323
Db 241 DNGSSEINLA-----MSGLORRKSRVRLSELGLGTYTKSGSNIRK----- 283
QY 324 EEVYVLRSSSKRRKTDKLLKKKOOH-SKRTAADVSD-AKLC----- 364
Db 284 EESALKKESEYGRK--RKLPENNYVSRILSTMGATSENAKSGCDSDQGNSESTDSGFR 341
QY 365 -----RRKPKVRLSEIINANOVEDSRDEVRHENAADCEDRST-----IP 408
Db 342 TPEKQORNRFOYVDEFEVSLPCEFSQ--EGAKEHADPSK--RSTPAHSLFTGNDVSP 397
QY 409 VPMEVSDIYVSNHTGEGDLKSSKNTKRYKDYVDVDSGLM---NWLNGKK--KRTG 462
Db 398 CP-----PGTQRT--ERKLSLPKKTKK--PVIDNGKSTVLSFGNGIDGSOVNSHTG 445
QY 463 SVHHTVAHPAGNLNKKVPTASTQHDNDENTENGLDTNMHKTQ--VCOHVEISTQRCS 520
Db 446 PSMNTVSQTDLLNGKRV-----GGLFDNRLASDGYFRKTLISOYNDKPT 490
QY 521 S-----KGTAGLSKGTHTSASTKYGESTRNGONIHVLSADOCOMETENSVLS 571
Db 491 SLHLQDNDYVRSDAEENCLRDPSSSSKSSGGWLRFG--VDIYDFNNHNTNRSSFSN 548
QY 572 HSAKVPSEHDIQMSDL-----HEOSLPKKKKOKLEVTREKOTMI 613
Db 549 LKRYRPPSETEVADLSVYLQKDSAGADRKGTAVVQVQDHGAPRSQSHREKETEEDONN- 607
QY 614 DDIPMDIVELANKNOHEROLM-TETDCSDINRISKTTAADDCYIVAKGSDVASSVFD 672
Db 608 DDIPMEIVEIMAKNOYERCLPKDEEDVSNKPOSETAHKSKNALIDLETNYDNGISLED 667
QY 673 TNSOQ--KSLASQSTOKELQGLHALTTOESPH-----PON--FOSTOEOQTHLR 717
Db 668 NNTSRPRKPCSSNARREE--HEPMGRQONSHPFISQPVVPFGIFPPTQENR---- 720
QY 718 MEKVNTIAASSPLFSHHDDQYIAEPTENHGRKAKKLTMEQFATRTNSPAATCGA--- 774
Db 721 -----ASSIRPSGNCMLGNLPTV--GNONPSSSFRVLNA-----CDTQOSVPM 764
QY 775 QFRPGIQAVALDSTHVMGSSSVYASROPVIAPLDRYAERAVNOVHARNPSTIATMEASK 834
Db 765 QYR-----EASHPTMPSS-----MTPPOQYRPVSLNTINQSTN-FGTLS--QASN 806
QY 835 LCDRRNAGVVLVYPK-----SMPTATLLRMM 861
Db 807 NEMTNWNLNFAVANGKQCGPNPEPSFCGKHAAGVSSSSSRPIDNFSSESSIPALHLSTL 866
QY 862 DPSTLASFP--NIGTS--SKNOMESQJHNSQVHHNRYKSTSTSYSGNLNKGITPLEEDL 917
Db 867 DPLRLSTTPADQHOHNTKFTKRHPNAPQSKFEIQLQGDSSKSAVSYTK---QIP--FDLY 921
QY 918 SRHQLDHLRPLRP-HRVGLGSLQKEIAMSSENGTOSGYLGVSTGITHQNRKE 976
Db 922 SKRPTQSPKSPFTYPTIGT--SLSFQNBKMSPH-----HQEKTK 962
QY 977 HFEALNSGMFSAKMNAIQLSVSSSAD-----FLSARNSIAOSWTRGKGKKNVHL----- 1026
Db 963 RKDP-----FAPVYNTHEKRPVFASSNDQAKRQLLAGANS-----MMPLKTKFHTMT 1006
QY 1027 DREVYRD-----ICITKNPADFTTISNDNEYM 1054
Db 1007 DKERKQKRKAESCNNNASAGPVKNSGPIYCSVNRNPADFTIPEPGVNYM 1056

RESULT 4
ID 09Y4D6 PRELIMINARY; PRT: 1983 AA.
AC 09Y4D6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)

```

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE KIA06524 PROTEIN (FRAGMENT).
GN KIA06524.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RT Nomura N., Ohtsuka O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL: AB014524; BAA31599.1; -.
FT NON_TER 1 1
SQ SEQUENCE 1983 AA; 221956 MW; EE1676A3B8ED37F9 CRC64;

Query Match	4.0%	Score 221	DB 4	Length 1983
Best Local Similarity	18.6%	Pred. No. 6.2e-05		
Matches 210	Conservative 111	Mismatches 423	Indels 328	Gaps 53

0Y	83	HDO---	KKCDEHKASSSPFVAEFRHMDSC	CLDKT	KTDNGAPARTLPKONG	SDCS	139						
		:	:	:	:	:							
Dd	947	HDEVDVAKCHSHSF	RNEBRGK	KIRH	-	ISCIEKLSKTESISVP	-----TSDHRS 995						
0Y	140	ITVVRSTFVPAVSQKVS	-----	PSOSS	---	QCKMADRSTLPKSVQEG	-----181						
Dd	996	L-----	IEANOSNSKVS	EDIT	YTCFLPKRSSSFL	FIHROSOSK	IMAAISLNGPPFOI 1048						
0Y	182	-----	NDSCNAPSGKNGAAE	NTSPK	ODLOGP	ONIDVAAN	SEDTSV 228						
Dd	1049	KNNEVDAMGN	MLNKFSPSPES	ANESK	YLSDBA	---	LEAPERTRMNVYSSOSTVR 1109						
0Y	229	VGALP	-----	EVPQITWHIE	-----	VNGADQPPSTPKLSE	YVLLKNEDENGKT 271						
Dd	1106	KGP	LPFLINRAMSCP	SEBP	HASTGREG	KKRLPG	SDADASELPPRAERILISEPDS-- 116						
0Y	272	EETLVAECCNLT	K-----	-----	-----	DPNP	-----MSGKERQ 296						
Dd	1164	-----	VRCOSILTKHQ	OKENPQ	ETIEKEG	MASSRVS	VALNSEDPLPCPSDLSKREK 1218						
0Y	297	VAEQCNLT	KDP	PVSGKCE	IOICNE	PEYVLLK	RSSSKRKTKDKIMKQOHSK----- 351						
Dd	1219	TLHKVK	-TTSTFVSQDE	---	DWVKCLEV	VSITYU	LLPFRKSPKFCNLLQOYTONTMLI 1273						
0Y	352	RTAODVADAT	LCKRRPK	-----	VLRLSEIT	IANOVEDSR	SDENVRENAADP 399						
Dd	1274	ESPQVET	FFNALKEK	OKONT	STREOS	GTSCENL	KAVNSDQTLTTEMKTAFRLSNGR 1333						
0Y	400	CEDD	-----	RSTIPV	PMEVSM	IPVSNHT	VGEDEGLKSSKNKTRKRYSDVVDGDSLMN 452						
Dd	1334	LAP	LOEMASVEAA	VSP	REESEK	AREIF	SDNLKATPLGJOSENKER----- 1379						
0Y	453	WLN	GKKRRKTS	VHHTVA	HAPGN	LSNKKV	UTPLASTQNDHDEDTENGLDTMNHKTIDVOCHVS 512						
Dd	1380	---	GKKLOSET	LHNSLM	LQKNV	SEER	-SENCQOSINSSNSGSSILPA	LSEVINIGNOST 1433					
0Y	513	EISTOR	CSSKGT	-----	AGLSK	RKTHSA	ASTYVG	-GESTRNGONIHVLSADOCOMET 565					
Dd	1435	RSS	SWECGSGRA	LPTG	SGCPC	OKDHT	STAVGDSGSGSOPRGRG	-----DGTGNQCKMT 1490					
0Y	566	ERSV	SHSAK	-----	VSPREH	IOIMS	LHE-----	OSLPK	-KKKOKILEVREKOT 611				
Dd	1491	-NKT	LISHSE	SOV	FAL	TPRL	HLQJ	GEETO	SDERPLSES	RELPFOR	SOEAMN	ETSK- 1544	
0Y	612	MID	IPMO	IVELLAK	NOHER	LMET	QCS	DIN	ROS	KTAD	DDCV	IVAAK	DSGY----- 666

```

Db 1549 -----AEDEMOMKSAIDQOSLPEGNK--NKTMLD-----LYKGEMNSVYKRLA 1590
QY 667 ----ASSVEDTN--SOOKSLAS-----OSTKCELOGHALTTOE-----SPHPONQS---- 708
Db 1591 AMSKASRKFPAKDVPSPRRHVAITFPQSGSRSGDFLSLSTVCNCLFPEPPKSAESIGE 1650
QY 709 --TQEOOTHLMEE-----MYLAASSPLFESHDOY--IAAPAEHNGR 749
Db 1651 SRUSENKHVKKSENLPTVLVLPNREPSTHVSQKSNSTISQHQMEFKNVSSPSKHEHS 1710
QY 750 KDAKLITWEOFKATTRNSPATGAOFRPGIAVDLTSTHVGSSSNVYASROPIAPLDR 809
Db 1711 KD-----VTAQNLVREGAP-----SPITFSLREAFESDNORLSPPF--PLEP 1754
QY 810 YAEAAVQVYHARNPSTIAIMESKICLDRNAGQVLYLKESMPATHLLRMDDPTSLAF 869
Db 1755 -AOK-----SKVSSPLASFQOORSASTL-----EVEPEPHLYR--SKLSKSI 1794
QY 870 PNYCTSSRNOMESOLHNSQYAHNOYKGTSTSGYNLCKIPLEFEDLSRHLDHLARPL 929
Db 1795 NVHGDLLRKS-----HPPKREHNFSESTSI--DNALSRLTLGNFVSNNVYSRRFPFSF 1846
QY 930 RPHRPRVGLSLLQKETIANSNENCFOSGYKLGAVSTGITSHOMNRKEHFEALNSGME$AK 989
Db 1847 SELPSCD--GN-----ESWAVYRSGKTKTGSRASISI-----YRPIDYGI$EKE 1886
QY 990 WNAQLQTSVSSADPLTSARNSIASOSTRGCKGVHDLRFVRODICITKNP 1041
Db 1887 QQLAFELNVRKS--LTQGLMKP$FLKPNQ-----FLKDDL-----RNP 1923

```

RESULT	5	
Q9LH98		
ID	Q9LH98	PRELIMINARY;
AC	Q9LH98;	PRT; 2081 AA.
DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DT	01-OCT-2000 (TrEMBLrel. 15, last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 15, last annotation update)	
DE	GENOMIC DNA, CHROMOSOME 3, BAC CLONE: T19N8.	
OS	Arabidopsis thaliana (Mouse-ear cress);	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_Taxid=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=COLUMBIA.	
RA	Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;	
RL	submitted (MAY-2000) to the EMBL/Genbank/DBD databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=COLUMBIA.	
RX	MEDLINE=20363099; Pubmed=10907853;	
RA	Nakamura Y.;	
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. II.	
RT	Sequence features of the regions of 4,251,695 bp covered by ninety P1	
RL	TAC and BAC clones.";	
RNA	DNA Res. 7:217-221(2000).	
DR	EMBL; AF002057; BAB03174.1;	
SO	SEQUENCE 2081 AA; 232851 MW; D3603E1F85EEFF29 CRC64;	

[illegible]


```

OY 362 KLCRRKPKVRL-----SEITANQ---VEDSRDEVHREN 395
DB 2243 SVPTSPKSTPTVEESTEQPTSTPGOSLTPMNSSEVLTTSEPHVLSLSPDVOSQS 2302
OY 396 AADPCEDDRSTTPVPMESMDIPVSNH-----TVGDDGKSSKNKTKRKYSDVDD 446
DB 2303 TTPNMLSESSVTEPP-KTSSSEVSLNSEPSTTEAPLTPLSPDLSTTNNLSQSTVSTED 2361
OY 447 GSSLNMLNGKKKRTGS---VHTVAPAGNLNKKKVTPTASTQHD--ENDTENGIDTN 501
DB 2362 RSEIS--ENSEKPTSAELVTSVTHVA---SSPDVPTSESPDDLTGSSTEN----- 2411
OY 502 MKTDCVCHVSEISTQROSSKGTAGLSKGTHTSASTRKYGESTRNGONHVLAEQOC 561
DB 2412 -----IPASSKQITSTPTPTDTTASEEPTKTSMPD----- 2445
OY 562 OMETENSVLSHS-----AKVSPAHDIOIMSDLHESLPKKKKOKLEVTREKOTM 612
DB 2446 -LSTSNVLSSESTPSSSKSPVSSSTEGISVSTEFSSVPESTLSSVLEEDLTKT- 2503
OY 613 IDIDPMOIVELLAKNOHEKQMTETDCSDINLQKSTADDCVIAAKDGSVYAFD 672
DB 2504 ---PSPLEETTTAETSEPLETDSLTYSVKRIHELITSEN--VPKESSTTSSSS 2556
OY 673 TNSQ-----KSLASOSTOKEL-----QGHALVT-----QSPHPQ 704
DB 2557 KPQOEAGILTSTVVVPTSSVSLTASEIEALTSMTPKQGRTPITTSKSLVKSTSPS 2616
OY 705 NFOSTQOQTHLMEMVTIAASSPLFSHHDOYIAEAPEHMGKDKAKLMEQFK--- 761
DB 2617 TVTSSSESESTKRTVTSTVSTTT-----PTEETTTSESLITLAAKSKPTE 2662
OY 762 ATTRNSPAATCGAO-----FRPGIQAVDLTSTHVMS-----SSNVASRQPYI 804
DB 2663 STTSSSAPTPPAKTSSTKRSNVSTSRKSTENVE--TSTOSGSLSSSTMSSTSEPETN 2721
OY 805 APIDRYAERAVNOYHARN-----FPSTIAMESKICDR----- 838
DB 2722 APAVTSSSEASTLEENSTSTSPSSSEASVKLSLEPEST--TSEAVTYSRPAITMS 2780
OY 839 ---RNGOVVLYPKESMPATHLLRMMDPS--TLASFPNYGTSSRNQMSQILNSOY--- 889
DB 2781 SESREISTVSESPSE--PEPLSTTVSPNVYASSIP-----SEBPLISSVYSSSTPRV 2833
OY 890 -----AH--NQYKGTSTSYGSLNKGKILTFEDLSRHQ----- 921
DB 2834 RLITGPDDLIVSYTVSHGNRRONITASSVPSNSTSPILLPSLTPPTPPPTTTTAK 2893
OY 922 -----LHDLHRPLRPHPRVGVLSLLQKEIAMSENCQSGYKLGIVST 965
DB 2894 PATTSYGKAPPSIOPPAPMFTTPAPPPSNNGYGETNOEBOYVSTTTTEAP---SLCS 2950
OY 966 GTTSHQNRKHEFEALNSGMSAKMNAIQLG---SYSSASDFLS-----ARNSIASWT 1016
DB 2951 TYTCHSLATCEQ---STGYCICRDGFIGDGTACSKSKSTADCLISPLCADKAKCDNST 3006
OY 1017 R-----GKGKV--HPIDRFYRQDIC 1035
DB 3007 RSCCEDAGYIGDYVCSPPQDCVLRDNLG 3036

```

```

RESULT 7
O44929 PRELIMINARY: PRT: 1690 AA.
AC 044929:
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE MICROBULE BINDING PROTEIN D-CLIP-190.
GN CLIP-190 OR CG5020.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OY NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=OREGON R;
RA Lantz V.A., Miller K.G.;
RL J. Cell Biol. 0:0-0(1998);
DR EMBL; AF041382; AAB96783.1;
DR Flybase; FBgn0020503; CLIP-190.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP Gly; 2.
SQ SEQUENCE 1690 AA; 189103 MW; BEAF48BD15F17A7C CRC64;

```

Query Match 3.9%; Score 215; DB 5; Length 1690;

Best Local Similarity 19.1%; Pred. No. 0.00012; Matches 171; Conservative 139; Mismatches 352; Indels 234; Gaps 33;

```

OY 174 LPSVOE-----GNSKCNAPSGKNGAAEANTDSFM---KDLQCPAQNYDVAA 218
DB 401 LQKINDELAKRIYELESALGNERK-----KTEELQCSIDEAQFCDELMANOSQYKREKI 454
OY 219 NVEDNTSYDVGLPEVPOITWHIEVNGADQPPSPKLSSEVVLKKNEDENGKTEETLVAE 278
DB 455 HDLESKITKLVSATPSLQSTL-----PEDLP-----SDDGALQDEIAGL 493
OY 279 QCNLTQPNPMSGKERQVNAQCNLTQDKPVPYSGQCEQIC-----NEPCEVYIK 329
DB 494 QEKWTIOQKEVESRIAGLEBEORLRENKYLN---EDIALQSELVSKDEALKEFSL 549
OY 330 RSSKSKRRKDKLKKQOHSKRTQAQADVSDAKLCRRKRYKYLSEITANQVEDSRD 389
DB 550 EGIENLRRELALKEENKQAOEAQEAFT--RLAKSVSVLSELDONLAKTSLSLES 608
OY 390 EYHRENADPCEDDRSTTPVPMESM--DIPVSNHTVGEGLKSKKRTKRYSDVDGCS 448
DB 609 E--RVNKSDECE-----ILQTEVRMDEQIRELNQDLDEVTQLNVOKA-----DSS 653
OY 449 SLMMNLNGKKKRTGSVHTVHAHPAGNLNKKKVPPTASTQHDNDENDENGIDTMHKTQVC 508
DB 654 ALDQMLRLQKEGT-----ERKSTLEKTEKELVQIKQAKKTLDQKEOLE 698
OY 509 QHVESISTQROSSKGTAGLSKGTSHAS--TKYGESESTNGONIHVLSAEDQOMETEN 567
DB 699 KQISDLQLAQEK-----LVREKTEANAIQLEKESISQOLAKONELEDEQKQKSES 753
OY 568 SVLSHSAKVSPAENDIQM-----SDLHQSLPKKKKKQKL--EYTRKQMTMDI 616
DB 754 EYHLEIKRQNTQKDLLEVESGESLKLQDLQLEKTLGHEHLQALDELKKEKETTIIKEK 813
OY 617 PMDIYELLAKNOHEKQMTETDCSDINRQKSTTADDCVIAAKDGSVYAFVDTNSQ 676
DB 814 EQLQLOQSKSAESASLKVVOY--QLEQLQOAAA-----SDEGSKYAKLHDEISO 865
OY 677 OKSLASQSTQKELQHLALTQESPHQNFQ-----STQOQTHLMEE 720
DB 866 LKSGA--EETQSELK-----STESNLEAKSKOLEANGSLIEBAKKSQGLQDQITKLSEV 919
OY 721 MYTIAASSPLFSHHDOYIAEAPEHMGKDKAKLTLWQPFATTTNSPAATCGAQRPGI 780
DB 920 EETQAA---LSSYHTD-----VESKTKQLEAANAAL----- 948
OY 781 QAVDLTSTHVMSSSNYASROPVIAIDRYAERAVNOYH-----RNPPSTIATWEASKL 835
DB 949 -----KNKEYASRAEASLDQKVEIYDITLHAELOARSSSALHT--KLSKF 996
OY 836 CDRRNGOVVLYPKESMPATHLLRMMDPSTLASFPNYGTSSRNQMSQILNSQVYAHNQYK 895
DB 997 SDELAICHKRLTSGKAAMSGEMQLK-----EKEQLDELROQLOQSDQSTYK 1043
OY 896 GSTSTSYGSLNKGKILTFEDLSRHQDLHLRPHPRVGVLSLLQKEIA-----NMS 950

```



```

Db 146 ANDFOQATIDOLENTPST-----PLRSOTVGDSTSTPAAPSLRKPIPISEV 194
Qy 199 NNDSPMKDLOGAON-----YD-----VAANVEDNTSV-----VGALEPVOITWHI 242
Db 195 VOSTVESVOSTSOHEKRIEVSLEYEFEMLEKNIHDEVNLOLKITELLKEKROTL-V 253
Qy 243 EVNGA---DQPPSTPKLSEVLK-RNEBNGKTEETLVAE-----QCULTDPNPMSCKE 293
Db 254 EIKEAREVEEKKIMOOVEAVKTKSERAAATLEOLLEKRIELEMKLIT-EPN-----GE 308
Qy 294 ROOVAECULTKDPKVSOGCEOICNEPCEEVULKRSK--SKRTDKMLKKQOOSKK 351
Db 309 KLOFEBHOLELK-----SRCEBLDKALKYDVOMQHSIEDYEKFEVLQEKKEADQL 361
Qy 352 RTQADVDSPAKLCRRKPKVRLIS--EITANOVEDRS-----DEVAR-ENA 396
Db 362 OKAKEDIETLOM-----KYVELETTINKVEFNSSEIETLSHEIYKRLMLDEIHRLENE 416
Qy 397 ADPCEDDRSTIPY-PMEVSM-----IPVSNHTVGEDGLKSS-----KNKTKRKYSDVVDGGS 448
Db 417 MSALQPRNDTTELELOKITDLKIDCCNLTKMLELOSEIVEVEKATSEIGEAVOKNG 476
Qy 449 SLMMNLGKKKRTGSV-----HHVYAHPRAGNLSNKKVPTASTOH-----D 490
Db 477 ELLEQJINSLHVENAKLIDMEGOLNDAHRKEDVDRISELTTTESLRDSEASDKLMD 536
Qy 491 ENTFENLDJNNMKTDCQHVSEISTQRCSSKGTAGLSKGTASASTRYGEGESTRNQ 550
Db 537 SESOTNEYSLALENT-----VSELETMRREYKA----- 564
Qy 551 NIHLVSAEDOCOMETENSVLSHSAKYSAPAHDIQIMS-DLHEOSLKRKKKKQ--KLEVTR 607
Db 565 ----SVDKVCSIQLELEIQTQHESTVELEAEIRIKELLELAOEAVKTVGSSOLKKEIYO 619
Qy 608 EK-QTMDIDIPMDIVELL-----AKNOHERO-----LMTETDCSDIN-- 643
Db 620 EDCOKLRQJLKEEQIOULVSLRETSVEMHOSAPAHQEKQIOJSLKMTSTAEVIELRSSI 679
Qy 644 ----RISQKTTADDCVI-----VAARDGSDVASS--VFDTNSQOKSL- 680
Db 660 DSLQAEVAVVQSDASADKHILEDLRIKTRQAEETNEKIRSLASSEQOIIDLKKNOESLI 739
Qy 661 ----ASOSTQKELOGHIALLTQESPHQNPONFOSQOQOOTHLMEMVNTIAASSPLFS 732
Db 740 DDLKEKLSHSESTQELQVLSLEMLKIEVSAR--QKYMSEVLEKESPEALOLEISA---S 794
Qy 733 HHDDQYIAEPTEHNGRDKAKITWEQFKATTTRNSPATCGAOPRPGIOAVDLSTHVG 792
Db 795 QEVSNSVYDANQ---EKDGLRLVYDLTKLKIETDEKS-----AODLOOSSV-- 838
Qy 793 SSSNVSAROPVIAPIADRAERAVNOYHARNPSTIATMEASKLCDRRNAGVVLVYKRESM 852
Db 839 ----SSSNVSAROPVIAPIADRAERAVNOYHARNPSTIATMEASKLCDRRNAGVVLVYKRESM 852
Qy 853 PATHLRLMMDPSTIASFRYGTSSRNOMESOLHNSOYAHNOYKGTSTSYSGNNGKTIPL 912
Db 866 NSSH---KRDYVALAS-----QLELOHKLVGSGYEVANKKEELIGAKI----- 906
Qy 913 TFEEDLSRHOLHDLRPLRPHRYGVIGSLQKEIAMSNGCTOGYKLGVSIGTISHOM 972
Db 907 ----MKKEWDELIN-----AKLGDALE-----GME--ELKKSLEYSEAKV 940
Qy 973 NRKEH-----FEALNSGMSAKMNAIOLGVSASSADFLSARNSIAQ 1013
Db 941 QREEEELIAOVSKHNDQOEOLOLTDELKSAQHSSTSTSSQSNELAAAREELIASISFQ 1000
Qy 1014 SPTRGKGMVHPDLRFVRODICTITKNPA-----DETT--ISNDNEY 1053
Db 1001 KALQDVEDVXH-----QODIQISEANEMAVVCLKODFERTSTLONEF 1042

```

RESULT 10

```

09QJ16
ID 09QJ16 PRELIMINARY; PRT: 1520 AA.
AC 09QJ16:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2001 (TREMBLrel. 19, Last annotation update)
DE 086.
GN 086.
OS Human herpesvirus 6B.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_Taxid=32604;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=229;
RX MEDLINE=99412318; PubMed=10482553;
RA Dominguez G., Dambaugh T.R., Stamey F.R., Demhurst S., Inoue N.,
RA Pellett P.E.;
RT "Human herpesvirus 6B genome sequence: coding content and comparison
RL U. Virol. 73:8040-8052(1999).
SR EMBL; AF157706; AAD49674.1;
SQ SEQUENCE 1520 AA; 171362 MW; 0F49EBEDCDF70F3D CRC64;

Query Match 3.8%; Score 207.5; DB 12; Length 1520;
Best Local Similarity 18.6%; Pred. No. 0.00029;
Matches 200; Conservative 170; Mismatches 480; Indels 227; Gaps 41;

Qy 84 DOKCKDEKASSPPSVAKFRFRMDCSKLDKL-KTSDNGTAPRTLPK--QNGTSDGS 139
Db 270 DNKEHKQOVSTSPALSLSTR-PSSQMDLIKITYTEVIR-LPEALIDNGSYCGVD 326
Qy 140 IFVSTFVPAVSVGSQKSPSTOSSQGNADRSLPK-SVQEGDSCNAPSGKNGAAEA 198
Db 327 TQKYYKTHRCK-----SIQTKGHS--QTINKKKYQKNEN--HVPSS-- 366
Qy 199 NNDSPMKDLOGAONDYAANVEDNTSV-----VGALEPVOITWHIEVNGADQPPSTPKL 256
Db 367 --RSLDKQRKS-NOHEDAVTEARDFSKLDPLSLPLMTPP-----PTL 407
Qy 257 SEVYLK--RNEBNGKTEETLVAEOCNLTQDPNPMSCGERDOVAEQCNLTG--DKRV 310
Db 408 NPAVHKTKHSSELHNTKKNHRSKTSL--QDVLISKHAPRAPTKNSYKKNHHDPT 465
Qy 311 SGOKEOICNEPCEEVYLKRSKSKRKTDKILMKQOOSKRTAQADVDAKLCRRPKK 370
Db 466 NDKP-----MKHSRGRTTSKKNTPANSNGHODVREAVVKNVSGKAANKTSK 511
Qy 371 VRLSELINANOVEDSRDEVHRENAADPCEDDRSTIIPVMEVSMIDIPVSNHTVGEDGLK 430
Db 512 ----SD--PSHNLHGKTSDBOYKTS-----PDNEKISTPRKSKTHHCIDHSSSEGOY 559
Qy 431 SSRNKTTRKYSDVVDGSSILMNLGKKKRTGSVHHTVANHPRAGN-LSNK-----KYPT 483
Db 560 KSPNNSSESNY-----GNCLSDKFFYNPFKTIPTN 587
Qy 484 ASQOHDENDENG-----LDTNMKRTDV-----CQHVS 512
Db 588 SKTNHKTSEESTENTDLNFSNENNTKTEIEDSNIIQPFQOLCNETIIPSTACPTQETP 647
Qy 513 EISTORCSKGTAGLSKG-----KTHS-----AATKYGGESTRNGVINHLASPD 559
Db 648 STMNRCASKSHSGANKNLTDNSPISSHNSPSTAFKNSNGSTNMS-----TSNGD 702
Qy 560 QCOMETENSVLSHSAKYSAPAHDIQIIMSDLHEOSLPRKKKKQKLEVYTRKQTMIDIDIPMD 619
Db 703 ECTDKRPN-----CSTENKSETSNOTNGENSDKPLSKTTEVDAASSRASSRASS 754
Qy 620 IVELLAKNOEROLMTEDCSDIRIQSK--TTADDCVIAAKGSDYASSVFTDINSQ 677
Db 755 RASSRASSRASSRASSRASSRTSSRASSRAFRASSRASSRASSRASSRASSRASS 814

```

QY 678 KSLASQSTOKELQHLALVTTQSPHPNOFQSTQ-ECQTHLRMEEMVTTIASPLFESHDD 736
DB 815 RASGRASGRASSRASSRVSRRASSRASSRASSRASSRASSRASSRASSRASSRASS 874
QY 737 OYIAPPEHMGKDAKLLTWEQFATRRNSPATCGAOFRRPGIADVLSTHVGSSSN 796
DB 875 RASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASS 934
QY 797 YASRQVPLAPLDYVABRAVNOVHANFPSTITAMEA-SKLCORRNAGQVY-----LTP 848
DB 935 RASGRASSRASSRASSRASSRSGKTPNNKLMNSIPSPQTYETKTSKRRAPQIYCDNKRQIYP 994
QY 849 KESMPATHLLRMADPSTLASFPNYGTSRRNOMESQLHNSQVYAHNOY-----KGSTST 900
DB 995 HDTSTSTEEVSEIKFRGCPPLNTFYKMAARLQSFNHNQOFYNPRFRPIRTNKKSEST 1054
QY 901 SYGSLNKGKIPILTFPDLRHHQHLHRLPRPRVGLGSLQKELTAMWSENGTQSGYK 960
DB 1055 NYTQDESS--TSRSRSHSHSPDLSINTPKRKKHKG--SSSISSSI--EENSRSMSRTE 1107
QY 961 LGVSTGTF---SHQNRKHEHFAALNSGMSAKMNLQLSVSSAD--FLSARNSTIAQ-- 1013
DB 1108 TGTDLTLTFNQHSTRSS 1167
QY 1014 -----SWTRGKGMVHP-----LDRFVRODICITNKNPAPDEFTTISNDREMYDY 1056
DB 1168 KOTRKSASRODKTKIKSPHESKHNHADWF-RNSQKTGKFLDLSNSPNTHEQSNH 1223
RESULT 11
P91257 PRELIMINARY; PRT: 3484 AA.
AC P91257:
DT 01-MAY-1997 (TREMblrel. 03. Created)
DT 01-OCT-2001 (TREMblrel. 18. Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19. Last annotation update)
DE HYDROTHERICAL 385.7 KDA PROTEIN.
GN P12F3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; Pubmed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton B., Wohlmann P.;
RT "The sequence of C. elegans cosmid p12f3.";
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: U80022; AAC25885.2; -
DR InterPro: IPR001064; Crystalin.
DR InterPro: IPR003962; FNIII_repeat.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003599; I9.
DR InterPro: IPR003598; I9.
DR InterPro: IPR003600; I9_c2.
DR InterPro: IPR003006; I9_MHC.
DR InterPro: IPR000130; zn_MTpeptidse.
DR Pfam: PF00041; fn3; 7.
DR Pfam: PF00047; I9; 10.

DR PRINTS: PR00014; ENTRYEIII.
DR SMART: SM00060; FN3; 8.
DR SMART: SM00409; IG; 12.
DR SMART: SM00408; IGc2; 8.
DR SMART: SM00410; IGc1; 1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
KM Hypothetical protein; Immunoglobulin domain. Repeat.
SQ SEQUENCE 3484 AA; 385746 MW; 21DD6F6893020F4F CRC64;
Query Match 3.7%; Score 206; DB 5; Length 3484;
Best Local Similarity 18.5%; Pred. No. 0.0011;
Matches 214; Conservative 161; Mismatches 419; Indels 360; Gaps 46;
QY 31 VLELATPPODAAEAGVDEPRAHOCENHSINGYVALLO--KKDP--KFCSLSTIFPD 84
DB 368 IIRKRETPESQVTA--EPEQKISEVDVOS-VAETEVGAKKPKDAEKPTDLSAKKD 423
QY 85 --QKQDEHKASSSPSVAKFRMDSKCLDKLTKTSNDCTAPRTLPKONGT----- 134
DB 424 SKSKSDPEASTEEKSTTEKPTND-----KTSKSAEKKTVAPKKEVTGKPLEAKK 475
QY 135 -----SDGCSITFY-RSTFVPASVGSQKVPSTOSSQ----- 165
DB 476 PVEDKKDAQSPSSKSSPPTDGKKKKQIIPKALFIPDELSHSRGDPTMHSFNITTTIR 535
QY 166 GK--NADRST-----LPKV-----QEGNDSKCNAPSG----- 191
DB 536 GREGSADAKTPLVEPLASVSKVESAKAEFSFRSRSETPDKSRKKEGLPPAKKSEK 595
QY 192 -----KNGCAEANTDPSPKDL-----QGPA-QNYDVAANYSE-----DNTSVDGALP 233
DB 596 KDEVTAKEQSTALIESKKEVEDESKISQOPSQDKKSEVGVPEKAAPEIKKDYSEIE 655
QY 234 EYP-QITWHIEVNGADQPSSTPKLSEVLKRNEDENGKT-----EETLVAEQC 280
DB 656 EYPAKKTKIKKTEKSD--SSISQKSNVLPAPDDKSKSDVDYDKSKKTTEQDTKVATDS 712
QY 281 NLTQPNPMSGKEROVAYQCNLTQPKPVYSGOKCQOJCN-----EPCEE 325
DB 713 KLEKADTTQJLETETVVD--KSKKRVLLKRTKESDSFISQKSETPPVVEPTPPAS 768
QY 326 VYLV-----SSSKSKRTDKLMMKKOOSKKTAAQ-----ADVSADKLCKRK 367
DB 769 EAQKIAEVNKAQKQKEVDNLRKAEVAAKKTADKLTAEANIKTKTAEVAAKQKQK 828
QY 368 PKRVLLSEIINANQVED-----SKSDEVHRE--NAADPCEDRSTIIPVM 411
DB 829 DQQLKLETEVVSQKSAEKLLEKQAIKKAADAVKQKQKLENNKLEAKKSAADKL 888
QY 412 EVSMQIPVSNHVVGEGLKSSKKN-----TKRYSVNVYDGGSSLMNMLNG 456
DB 889 KLEESAASKVSESVFGEKKTKAGEKTVQVESEPTSKTIDTKVAGAEPADETP 948
QY 457 KR-----KRTGSVHHVVAHPAGNLSNKK-----VTPTAQSTOHDDND 493
DB 949 KKKIKKTEKGDSSISQKSAIDSEKVKQKQKEDDEPTKRAVSTQWYTEADSKKQKEND 1008
QY 494 TENGIDTNN-----HKTDCVQHVSEISTORCSSKGTAGLSKGTSAASTKYG 542
DB 1009 EKLKIDAEIAAATKQBADEKSKLDAQEIKKVSQEDDAARKEKLN-DKLTLESEIATTKA 1067
QY 543 G-----ESTNGQNIHLSA-----EDQCOMETENVLSHSAVSPAENHIQIMSDL 589
DB 1068 SADKLTLEQQAQKKAAYEAQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1127
QY 590 HEGSLPKKK--KKQKLEVTRE-----KQTMIDIPMDIVELAKNHOE 630
DB 1128 GADAVAKQKQKLEKKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNL 1185
QY 631 KQMTETDCSDINRIOSKTTADD-----DCVIYAAKQSGSD 665

```

Db 1186 KOTKLEDEKSTESSEKTEVDEKPKKVKLKRTKESDSSISQKSETSKTYVESAGPSES 1245
OY 666 YASSVFPTNSQK-----SLASSTOKELQ 691
Db 1246 ETOKVAADARKQKETDEKQKLEAITAKKSADKSKLEAKAEVKAARAKKOKEKDE 1305
OY 692 HALTTOESPHPONFOTO-EOQTILRMEBMTIASSPLFSHHDOYIAEAPTEHWGRK 750
Db 1306 QLKIDTEASKKAAAEKLEKQSHIKKAEVD-AVKKQKLEKORLSEATK---KA 1361
OY 751 DAKITMQEFATIRNSPATNG-----AOFPGICAVDLSTHWGSSSNVYS 799
Db 1362 DAERKJLEEOK-----KKAETILIEIQKQEKIAQOQSRLDEAKKSAKQKLESETKS 1416
OY 800 ROPVIAPIADRYAERAVNOV-----HARNFPESTI--ATMEAS-KICDRRNA 841
Db 1417 KOTEPAPKESVDEKPKKKVKLKRTKESDSSISQKSKASTVDAAEFTLESDFNLVEKKTY 1476
OY 842 GQVLYPKESMPAT 855
Db 1477 QKVEQSPDESTSAT 1490

```

RESULT 12
09N435 PRELIMINARY: PRT: 2083 AA.

```

AC 09N435;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOHETICAL 231.2 KDA PROTEIN.
GN Y38B5A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1996).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC Sun H., Geisel C.;
RA "The sequence of C. elegans cosmid Y38B5A.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC Waterston R.;
RA "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DL EMBL: AC024760; AAF59460.2;
DR InterPro: IPR003962; ENITI_repeat.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003959; I9.
DR Pfam: PF00041; fn3; 2.
DR PRINTS: PRO0014; FNTYPEIII.
DR SMART: SM00060; FN3; 3.
DR SMART: SM00409; IG; 3.
DR Hypothetical protein: Repeat.
KW SHOUENCE 2083 AA. 231174 MW; 9270303ADEZD7C2F CRC64;
SQ

```

Query Match 3.7%; Score 205.5; DB 5; Length 2083;
Best Local Similarity 18.6%; Pred. No. 0.0006;
Matches 175; Conservative 135; Mismatches 435; Indels 197; Gaps 30;

OY 1 MEIYAVDQEGARVGTNCMLARGTGAVAPVLELTATP----- 38

```

Db 467 IELVPEEFAFRIVAVNGF-----GEGAPSEILEVNTLDYQDESEDFAGEBELKIDVQV 522
OY 39 RODAAAGVDEPAHQCEHFSIRGVALLQKDPKFCSLIFHDOKCDEHKASS-P 97
Db 523 NNEVVEITIEESEVILIEH-----RLKRRSK-----SKTIDPELDSEIA 566
OY 98 FSAKFRMCKSCLOKRTSDNGTAPRTLPKAKNGTSDGCSITFVRSTVPASVGSQKV 157
Db 567 LEVS-----SDITSLTETTESITPDPAESQETLNEIAVT---ETTV-----OKI 610
OY 158 SPSTOSSGKNADRSTLPKSVQEGNDKSCNAPSGKGAALNMDSPKIDQAPQNYDA 217
Db 611 TNPDESAAKKNVNDTAVSSIVKXKDDVN---KSLPSGL-TKKELQKPEKKIMK 665
OY 218 ANVEDNTVDVGLALPEVPOITWHEVNGADOPPSPFKLSEVLAKRNEDENGKTEETLYA 277
Db 666 KTERKADDSISETS---EYTLKDLTQKSEPEPKARTTETSQV--DEVKRTETTSKS 719
OY 278 EOCNLTQDPNPMGKERDQVADQCNLTQKPKVSGQKE-QICNEPCEVYLKRSKSKR 336
Db 720 KO---TTEHQPQGGKSDSISSTSDASEVQVOOSEAKQVTEKPEYAKLSKSKMT 776
OY 337 KTDKILMKQKQSHKRTQADVDYDAKLCRRKPKVRLSETINANOVEDSRSDVHREN- 395
Db 777 DTTK-----ESDNKEIYDEKPKKKVKLKRTKESDSTISSETSAYESAGPSESTQNV 830
OY 396 AADPCEDDRSTIPVMEVSMIDIPVSNHT-----VGEGLKSKRKT-----436
Db 831 AAVDEKKQKENDKQKLEAELTAGKSTQKSKLEAEAKLRAAEADAAKQKQKTEAS 890
OY 437 -----KKRYSDVDDGSL--MMWLGKKK-----RTGSVHNT 467
Db 891 KKAAEKLEKQAIKKAADAVAKQNELEQKLEATKLAEEKLEQSAASKQ 950
OY 468 VAPHAGNLNKKYTPYASTQHDENDTENGDLTNNHKT-----DYQNHSEISTORCS 520
Db 951 AAEQAKILDQTKAKAAEKQTEGLEKDESKNKGSENEYEEKPKKVKLKRTKESDSSIS 1010
OY 521 SKGT-----AGLSGKTHSAASTRYGGESEPRNGONIH-----LSAEDQCOMETENS 568
Db 1011 QKSDTSKTVAESAGSSSESEFYADATSKOKETDRKQKLEAITAKKSDEKSKLETESK 1070
OY 569 VLHSKAVSPAENDIQIMSLDHSQLEPKKKKQKLEVTREKQTMID-----DIPMDIVEL 623
Db 1071 LITKAEDAAKQKQKEDKLEADVASKKAAAEKLEL--EKQAIKKAADAVAKQKLE 1128
OY 624 LAKNQHROQLTEPDCSDINRIQSKTTADDCYIVAAK-----DGSDYASVFTNSQK 678
Db 1129 AEKQKLESEAAATKKAAEKLEQAOINKAAEADAVAKQKQKLEKKNKLEANKSAEKL 1188
OY 679 SLASSTOKELQGHIALTTOESPHPONFOTOEOQTILRMEBMTIASSPLFSHHDOY 738
Db 1189 KLEESAAKSKQ---TVEQAKILDQTEKETAKEQOTGLENDKST-----KDSF 1234
OY 739 IAEAPTEHWGRKDAKKLWEOFKATTRNSPATGGAQFRGICQAVDLSTHWGSSSNVA 798
Db 1235 SKETVDEKPKKKVKLKRTKESDSSISQKS-----VTSKTVESGG---1274
OY 799 SROPVIAPIADRYAERAVNOVHARNPSTIATWEASKLCDRN 840
Db 1275 ---PSESEYOKVADAARQKETDEKQKLEAITAKKSADKS 1313

```

RESULT 13
076891 PRELIMINARY: PRT: 5327 AA.
AC 076891;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EG:49E4.1 PROTEIN.
GN FUTSCH OR EG:49E4.1 OR CG3064.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 RN (1)
 RA SEQUENCE FROM N.A.
 RP Papigannakis G., Spanos L., Siden-Kiamos I., Louis C.;
 RT "Sequencing the distal X chromosome of Drosophila melanogaster."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RA SEQUENCE FROM N.A.
 RP Benos P.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL031128; CAA20006.1; -
 DR FlyBase: FBgn0015390; futsch.
 SQ SEQUENCE 5327 AA; 575942 MW; FEFEE23A18FF33A CRC64;

Query Match 3.7%; Score 204.5; DB 5; Length 5327;
 Best Local Similarity 19.1%; Pred. No. 0.0025;
 Matches 225; Conservative 174; Mismatches 457; Indels 319; Gaps 51;

QY 31 VLELTATPPRODAAEAGVDEPAQHCE---HFSIRGYA---LQKKDKPKFCSLRIFH 83
 DB 1967 VAERSPLPEKESRPAVSVAESIKDEAEKESRESVAEKSPLPKESASRPAVSVAESIK 2026
 QY 84 DQ-KKODEH-----KASSPSPVAKFRRMCSKCLDLKLTSDNTARTLP 128
 DB 2027 DEHEKESKESRESVAEKSPLPKESASRPAVSVAESIKDEAEKESRESVAEKSPLP 2085
 QY 129 AKQNGTSDCSTITFVSTFVPAVG---SQKVPSTQSSQGNKADNSTLPKSVQEGNDK 185
 DB 2086 SKEASR-----PASVAESVDEADKSKESRESMAESGAGST-KQDQSP 2130
 QY 186 CNAPSGKCAAEANDSPKMDLQPAONTDVAANVEDNTVDVGLPEVPOITWHIEVN 245
 DB 2131 LKEVSRPESVAESVSKDDPKYSKE-PSRRSVAGSVTADSAKDQSPDLP-----ESK 2179
 QY 246 GADQPRS-TPKLSEVYLKKNEDENGKTEETIVAEQCNLTKDPMPMGKERDQVAEQCNLT 304
 DB 2180 GASRPSPVSDVSKDEAEKESRESSTESVT-----PPAKDKDSPEVQLQPSMT 2230
 QY 305 KDPKPSGCKCEQICNEPEEVLKRSKSKRTDKLMKQKQHSKRTQAQADVSAKLC 364
 DB 2231 ETIREADQPMKPSQASRESRESTIAESIKASPRDEKSPKASRPGSAESI---KTD 2287
 QY 365 RRRPKVRLSELITANQVEDSRSDVHRENAADPCEDDRSTIPVMEVSMIPV-SNHT 423
 DB 2288 LDKPQITK-----DDKSTHSRRESL-----EDKSAVTSKESVSRLPLVASDHE 2331
 QY 424 VG---EDGLKSS---KNKTKRRY-----SNVVDGGS---SLAMWLNG 456
 DB 2332 AAVAIEDDAKSSISPKDKRPGFVAETVSSPIEATMEFSKIEVKEKSLALSLOGSGG 2391
 QY 457 KKK-----FTGSVHTVTAHPAGNLSNKKVTPAS-----TQHDDENDTEGLPTNMH 503
 DB 2392 KIQTDSPPDVNAEGDSHAVA-----SVSTVPTLTLPKRAELAQIGAAKTVSSPDLBALR 2445
 QY 504 KTDVCOHVS-----EISTQRCSKGTAGLSKGTN-----SAASTKYGGESTRN 548
 DB 2446 TTSAPHERISRADSPACASEEIASQKSPQVLKESRPAWVAESKDDAQKLKSSVEDLRS 2505
 QY 549 GGNHILNAE---DOCOMETENSVLSHS---AKVSPAEH---DIQIMSDIHEOSLP- 555
 DB 2506 ---PVAETISRPAASGAEASSPIEAPRDFAEFQDAEAVLPLTITELKGNPLTSSPV 2561
 QY 596 -----KKKKKKOLEVREKOTMIDIPMDIYELLAKNOHEQLMTETPDS-----D 641
 DB 2562 DVNAHVAQVARELSKVIEKTASSPIDAKRSLIGSPAEEPRSPASAKDAAEVSKSD 2621
 QY 642 INF---IQSKTTAD-----DD----- 654

DB 2622 ASRPSPVESTKADSTYKGDISPSPESVLEBQPKDVEKESRHPSPVASITGDSTKDV 2681
 QY 655 ---CVYAAKDGSDYA-----SVFDT--NQOKSLAQSOYOKELGHALTTQ 699
 DB 2682 RPASVESVKEDEHDKAESRRESIAKVESYIDEGAKSDSKS-SSQDSQKDEKSTLA-SKEA 2739
 QY 700 SPHPNFOSTOEQOOTHLMEMEYVLTIAASPLFESHHDQYIAEAPTEHMKRDKAKLTWEQ 759
 DB 2740 SRRESVESKDKAESESRESPIASGEV-----PRESKSPDSKOTS--- 2784
 QY 760 FATTNRNPAATGGAQFRPGIAVDTLSTHVCGSSNVSAROPVIALDRYAERAVNOYH 819
 DB 2785 -----RGSVYESVTAE---DEKSEQSRRESVAESVADTKKKGKQ 2824
 QY 820 ARNPSTIATMEASKICDRNAGQVLYLPRESNPAH--LLRMDSTLASPNTYGTSSR 877
 DB 2825 EASRPSPVD-ELLKDDDEQOESR-----RQSTTGSKMASTYTGDDSPMDKADKESPS 2877
 QY 878 NQ--MESOLHNSQYAHNOYKSTSTSGSNLNG-----KIPLTFEDLSRHQLDHLRPL 929
 DB 2878 PESVAESIKHENKDESDPLGSRDSVAESIKDITKGEKSPKPSKEVSRPE----- 2929
 QY 930 RPHPRVGLSLQKEITANKSENCGTQSGYKLCVSTGITSIHMNRKEHFEALNSGMPSAK 989
 DB 2930 -----SVGSIKDEKAESRRES--VAESYKPPESKSDATSAAPS-REHSRPS----- 2973
 QY 990 WNALQGSVSSADPLSARN-SIAOSMTNKGKMY 1023
 DB 2974 ----VLGSLKDESDGKTTSRKVSVAADSIKDEKSLIV 3004

RESULT 14

ID 09VC00 PRELIMINARY; PRT: 2768 AA.

AC 09VC00; PRELIMINARY; PRT: 2768 AA.
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CG13648 PROTEIN.
 GN CG13648
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 RN (1)
 RA SEQUENCE FROM N.A.
 RP STRAIN-BERKELEY;
 RC MEDLINE-20196006; PubMed-10731132;
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrell J.F., Adyaeni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos P., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houson K.A., Howland T.J., Wei M.-H., Ibegam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Modarri C., Morris J., Moshrefi A.,

[illegible]

RESULT	15	PRT:	2951 AA.
Q9M3Z0		PRELIMINARY;	
Q9M3Z0			
DC	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)		
DE	CG3950 PROTEIN.		
GN	CG3950.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydriodes; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	SRRAIN-BERKELEY;		
PC	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sulton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,		
RA	Man K.H., Doyle C., Barker E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	April J.F., Advayani A., An H.-U., Andrews-Plamko C., Baldwin D.,		
RA	Bailew R.M., Basu A., Baxendale J., Brokstein P., Brothier P.,		
RA	Beeson R.Y., Benos P.V., Berman B.P., Brokstein P., Brothier P.,		
RA	Borkova D., Botchan M.R., Bouck J., Cadieu E., Center A., Chandra I.,		
RA	Buttis K.C., Cawley S., Dahlike A., Deng Z., Davenport L.B., Davies P.,		
RA	Cherry J.M., Delcher A., Deng Z., Mays A.-H., Dew I., Dietz S.M.,		
RA	de Pallos B., Delcher A., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dubin K.J., Evangelista C.C., Ferris C., Gelbart W.M., Glaser K.,		
RA	Foster C., Gabriellian A.E., Gary N.J., Guan P., Harris M.,		
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jatani M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Kethum R.A.,		
RA	Kimmel B.E., Kodla C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Mosnig A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Neuman D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,		
RA	Palatzkolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,		
RA	Shue B.C., Siden-Klamis I., Simpson M., Skupst M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Switsks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,		
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,		

RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003438: AAF6171.1: -.
DR FlyBase: FBgn0029875: CG3950.
SQ SEQUENCE 2951 AA; 332269 MW; 85CCCA0BBDAC8AEB CRC64;

Query Match 3.7%; Score 204; DB 5; Length 2951;
Best Local Similarity 20.1%; Pred. No. 0.0012;
Matches 213; Conservative 161; Mismatches 463; Indels 222; Gaps 48;

QY 78 LSRIFHDQKCDENKASSPFSVAFRRWDCS-----KCL-----DK-LKTSDNCT 122
DB 1838 VSRFTHSAEKVKEPKV-SAPFS---REFDVAKIPENDCLSLNQKDKFLRAENVI 1892
QY 123 APRTIPAK-----QNGTSDGCSITFVRSF--FVPASVGSQKVSPTSSQSGK- 167
DB 1893 RPTSLPSRPELEIRPELEIDEELRDCTLSVQVHKFIDF--AEKLAIPMPQKSPRL 1949
QY 168 --NADRSTLPKSVQGNDSKCNAPSGKNGAANPDSPKDLQGPAAQ---NYDVAAVSE 222
DB 1950 VANIRHISROSEPRELDESEF-----ELDRDTDYEDDDOTSQLETEEEITQYVTK 2002
QY 223 DNTSYDVGALPEVPQITWHIEVNGADQPEPTPKLSEVYLKRNEDENGKTEETLVAE-QCN 281
DB 2003 KET-----LKEFKQOTKETRETRRDSKAPEKLO---KSPQTKVKESARVPKYQAK 2052
QY 282 LTKDNPMSGKERDOVAEQCNLTQPKPVSQKCEQICNEPCEEVYLYKRSK-SKRKIDK 340
DB 2053 VSQKYSQWEPKKQPREPVTKETPLEPKKQPLSKVKDEP--EKVYKREPKVPQKESQT 2110
QY 341 KLMKQOQSKKRRKQADVSDAKLCRRKPKVRLSELIANQVSDRSDEVRENADPC 400
DB 2111 KKEPERYTKTTPQ-----KEPRKEPL-----ROSEDEPFSPEEEDDEPL 2153
QY 401 EDDRS-TTPVPEVSMDFPVSNHTVGE--DGLKSKNKTFRKYSDVDDGSSLMNLNGK 457
DB 2154 PWTKHTTAIEKKRQKDLNRPVSFGQRTPERKSSSTPSTKLNGTRGRSPSTNLITEE 2213
QY 458 KKRISGVHHTVAHPAGNLSNKKVTPASTQHDNDENDENGLDTNMHK-----T 505
DB 2214 KRSYRNOYVNWSP---GTRKTPSANSQAQSPPKTYSISKFMEOISQGSVVQDVV 2269
QY 506 DV-----COHVEISLTOGSSKGTAGLSKGTHS---AATKYGESTRNGQNIHL 555
DB 2270 DVEVVGPARPSISEKPOGKSPPTSSRLSPKSPKSTITTY---TTNTGRNVA 2325
QY 556 SAEDOCQMETENSVLSHSAKVPSEHDIOIMSDLHEOSLPKRRKKOKLEVTR---EKOT 611
DB 2326 SRRNVFEPVHEHTVDSEPTGRPSYMD-----HTKSLHHRDLSLEINKSHYSRKS 2378
QY 612 MTDIIPMDIVELLAKNQHROQLMTEGSDINRIOSKTTAAD--DCVIVAAKDSGY--- 666
DB 2379 MEDDSPVE-----PRNPNSVVF-----DVPKSSSSGADDEPRKTSLGKDESDLELE 2427
QY 667 ASSVFDTSNQOK---SLASQSTQKELQGLALJTQESPHONQSQEQOQTHLMEMVT 723
DB 2428 IEEIFDLORLEKLETVASIEKRRIRAOMLIR-----KNNINAGTTTTTTTITSTT 2481
QY 724 IAASSPLFSHHDQYIAEAPTEHMGKKDAKLWEOFKATTRNSPATCGAOFRPGICAV 783
DB 2482 PKSSPLPIRIRDOQPAGAA-----EYKTEKVTHTTSRROOQOQVDEOV 2524
QY 784 DLTSTHVGSSSNVASROPIAPLDRIAYERAVNOVHARNP-----STIATWEASK 834
DB 2525 DSTTPIAPGKTSPHG--KPPVAPRERSASPAQKR--RISPGKQSPGDRSTTTTKVYT 2579
QY 835 LCDRRNAGGVLYPKESMA-----THLR--MMDSPLASFPNVTSSRNQMESQL 884
DB 2580 TSTTKA-----PSKPAQGPIMADRSKVLGHAATVPQTINGSTPRKGSSTSS-- 2629
QY 885 HNSOYAHNOYKGSSTSYGSLNGKIPLEFEDLSRHQLHLRPLRPHRVGV---LGS 940

DB 2630 --GKITRTWSSSTTTSSSSTTNTNRKQREDSITSSY-----GVGPTDENGILPLFGIRA 2682
QY 941 LLOKELANMSEKCNQ---SGYKLG---VSTGTSQNNRKEHFALNSGMFSAKKNALO 994
DB 2683 LKKKATPPAEPECETKQEVYTGIVIEQFYSDNKSPPRHERKELIYSSNADELAIAIKOOLQ 2742
QY 995 LGSVSSADFLSARNSIAQSWTRGKGMVPLD-REVRQ 1032
DB 2743 -DEDDSSPPLDAR--VYREFFKVESQOSLPEDARYVR 2778

Search completed: September 16, 2002, 22:34:15
Job time: 654 sec

THIS PAGE BLANK (USE FOR)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2002, 20:09:51 : Search time 65.82 Seconds
(without alignments)
1783.728 Million cell updates/sec

Title: US-09-828-068-2

Perfect score: .5526

Sequence: 1 MEIVAVDQSGARVGTNCML.....NKNPADFTISNDNEYMDYR 1057

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A-Geneseq_032802:*

- 1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	393	7.1	1096	21	AAV84833
2	221	4.0	1982	22	ABG16404
3	221	4.0	1982	22	ABG19656
4	220	4.0	1325	22	ABB29725
5	220	4.0	1325	22	ABB34897
6	220	4.0	1325	22	ABB20314
7	220	4.0	1325	22	AAW55707
8	220	4.0	1325	22	AAW68085
9	220	4.0	1325	22	AAW05645
10	217	3.9	1165	22	AAW07091
11	208	3.8	1528	18	AAW20056

12	208	3.8	1583	18	AAW20057
13	206.5	3.7	2070	22	ABG03703
14	204	3.7	2768	22	ABB68397
15	204	3.7	2951	22	ABB60291
16	201.5	3.6	647	10	AAW93971
17	201.5	3.6	665	21	AAW18278
18	201.5	3.6	1784	22	AAW71460
19	201	3.6	2344	22	AAW37120
20	200	3.6	3111	22	ABB60327
21	198.5	3.6	1690	22	ABB61144
22	198.5	3.6	1690	22	ABB61173
23	198	3.6	1552	21	AAW85577
24	197.5	3.6	754	21	AAW70008
25	197	3.6	694	17	AAW85596
26	196.5	3.6	2515	22	ABB12281
27	196.5	3.6	2515	22	AAW80268
28	196.5	3.6	2515	22	AAW80268
29	196.5	3.6	2519	22	ABG16636
30	195	3.5	1730	22	ABB68370
31	194.5	3.5	673	11	AAW05701
32	193.5	3.5	1803	22	ABB65391
33	192.5	3.5	2025	22	AAW34207
34	192.5	3.5	3158	22	AAU37018
35	192	3.5	3201	22	ABB62899
36	191.5	3.5	982	12	AAW13320
37	191.5	3.5	2633	22	ABG06505
38	188.5	3.4	2897	22	ABB58514
39	188	3.4	2505	22	AAW40002
40	187.5	3.4	2586	22	ABB66878
41	187	3.4	1420	22	ABB63410
42	187	3.4	1489	22	ABB59948
43	187	3.4	1863	17	AAW81505
44	187	3.4	3257	22	ABB67502
45	184.5	3.3	2663	22	AAW39097

ALIGNMENTS

RESULT 1

AAV84833 standard; Protein; 1096 AA.

AAV84833;

08-AUG-2000 (first entry)

Amino acid sequence of the EMF-1 gene.

Embryonic flower gene-1; EMF-1; reproductive development; flowering;

early flowering; uniform flowering.

Arabidopsis thaliana.

WO200022132-A2.

20-APR-2000.

08-OCT-1999; 99WO-US23543.

09-OCT-1998; 98US-0169696.

(REGC) UNIV CALIFORNIA.

Sung ZR, Aubert D, Chen L;

WPI; 2000-317987/27.

N-PSDB; AAA14873.

An isolated nucleic acid molecule comprising embryonic flower genes 1s

useful for modulating reproductive development in plants

Claim 7; Page 34-38; 38pp; English.

expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1325 AA:

Query Match 4.0%; Score 220; DB 22; Length 1325;
 Best Local Similarity 19.4%; Pred. No. 9.7e-08;
 Matches 153; Conservative 125; Mismatches 311; Indels 199; Gaps 38;

QY 113 DLKLTSDNGTAPRTLPKAKONGTSDGCSITFVSTFVPASVSGSKVSPSTOSSGKNADRS 172
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 536 dtrplanektlpslaeptengq-----rtpfanekttssaeptehertplanen 586
 QY 173 TLPKSVQEGNDSCNAPSGKNGAENATDSPMKDLOGPANTDVANVSEDNTSDVYAL 232
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 587 ttp-spaepent-----ertanentlpsp---agptenremtan---ektlilfaep 632
 QY 233 PEVPOITWHIEVNGADQPS-----TPKLEVVYLRKRNEDENGKTEETIVAEOCNLTKDP- 286
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 633 tenertanekttspaeptengqrtpfanekttspaeptehertplane-ntlpspa 691
 QY 287 NPMSGKE-----RDQVAEOCNLTKDP-----KPVSGKCEIOCN 320
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 692 eptenreretanektltpfaeptenreretan-ntlpspgtengdrtplanektlpsla 750
 QY 321 EPCSEVYLRSSSKSKRTKTKLKKQHSKKRTAQADVADAKCRKKPKVRLSEITINA 380
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 751 epteng-krtpfanekttssaeptehaer-----rtplane 785
 QY 381 NQVEDSRSDVHRENA-----ADPCEDRS-----TIPVMEVSMIDIPVSNH--TY 424
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 786 ntlpspaepentertanektltpfaeptenrestanektltpfae-----plenrewta 840
 QY 425 GEDGLKSSKNKTKRKYSDDVDDGSSLMNMLNGKRRKRTGSHHTVAHPAGN-----LSNKK 479
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 841 nentlpspepteh-----eemtplan-----ektlis-----paepentgerptltnek 885
 QY 480 VWP7AS--TQHDENDTENGDLTMMHKTVCQHVSEISTQRC-----SSKCKTAGLSK 530
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 886 ttpsaeptehertplane-rtpsaeptehertplanekatpsakplehgett-vne 943
 QY 531 GRTHSASRTKYG-----ESTRNGONIHVLSAEDQCOMETENSVLSHSAK 575
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 944 dtcpsaeptengertplanenttlpsstestehertanektlpsaeptehertpsan 1003
 QY 576 -----VPAHEDIQIMSDLHESLPRKKKK-----QKLEVTREKQIMIDIMDIIVELL 624
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 1004 ektlpsakplethe-emptsanentpspykplehgektllanektlispeep-----tehg 1059
 QY 625 AKNQHERQIMLETDCSDIRIOSKTTADDCYIVAAGSDVYASVFDN----- 674
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 1060 akttsanektlpslakpleh-gerltspndklttsaestehdratsanvltpapaept 1118
 QY 675 --SQQKSLASQSTQKLOGLHALITQESPHPOFOSTOQOITLRMEEWNTIAASSPLRS 732
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 1119 khaektlishekmctq-----vtekstenpektsttektl-rtpekrplysektic 1169
 QY 733 HHDDOYIAPEPTHEGMRKDAKTLTWEQFKATTNN--SPATCGA--QFRP-----GICAV 783
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 1170 kgknlpvpekprenlgn-----tlttetlikapvksenpektaaavtktlpsvavtgdksl 1226
 QY 784 DLSTHV 790
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 1227 tltsshl 1233

RESULT 5

ID ABB34897 standard; Peptide; 1325 AA.

AC ABB34897;

DT 04-FEB-2002 (first entry)

DE Peptide #2403 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 27; SEQ ID NO 27532; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX fetal liver. The present sequence is a peptide encoded by a single exon

XX nucleic acid probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1325 AA:

Query Match 4.0%; Score 220; DB 22; Length 1325;
 Best Local Similarity 19.4%; Pred. No. 9.7e-08;
 Matches 153; Conservative 125; Mismatches 311; Indels 198; Gaps 38;

QY 113 DLKLTSDNGTAPRTLPKAKONGTSDGCSITFVSTFVPASVSGSKVSPSTOSSGKNADRS 172
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 536 dtrplanektlpslaeptengq-----rtpfanekttssaeptehertplanen 586
 QY 173 TLPKSVQEGNDSCNAPSGKNGAENATDSPMKDLOGPANTDVANVSEDNTSDVYAL 232
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 587 ttp-spaepent-----ertanentlpsp---agptenremtan---ektlilfaep 632
 QY 233 PEVPOITWHIEVNGADQPS-----TPKLEVVYLRKRNEDENGKTEETIVAEOCNLTKDP- 286
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 633 tenertanekttspaeptengqrtpfanekttspaeptehertplane-ntlpspa 691
 QY 287 NPMSGKE-----RDQVAEOCNLTKDP-----KPVSGKCEIOCN 320
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 Db 692 eptenreretanektltpfaeptenreretan-ntlpspagtengdrtplanektlpsla 750

Qy	321	EPCEGVVLAQSSSSKKRKTDDKKLKKQOHSKKRRADVDSPAKICRRKKPKVRLLSITNA	380
Db	751	epfeng--krtflfnektltsssaeptheaer-----lplane	785
Qy	381	NOVDSRSDEVHRENA-----ADPCDDRS-----TIPYMEVMSMDIPVSNH--TV	424
Db	786	ntspaeptentartanektitfpaeptentrestanekektlfpae-----ptenrewta	840
Qy	425	GEDGLAKSSKNTKKKYDYVDGSSLLMMNLGKKKKRTGSHHTVAPAGN-----LSNKK	479
Db	841	nenltlspaepteh-----eemtrplan-----ektls---paepfengertpfinek	885
Qy	480	VTPPAS--TOHDDENDRENGLDITPMHKTDCOHVSEISTQRC-----SSKGTAGLSK	530
Db	886	tlpsaepfengeitrlanel--tlpsaepfengerianekakrpspraktehgett--vne	943
Qy	531	GKTHSAASTKYGG-----ESTRNGQNIHVLASBDOCOMETENSVLSHSAK	575
Db	944	dtpssaepfengertplamenttspstestehgerianektlpsaepfengeitrlpsan	1003
Qy	576	-----VSPAEHIQIUMSLDEQSLPKKKKK-----QKLEVTREKQIMDDIMDIYELL	624
Db	1004	ektlpspraktehe--emtpsanenttlpsrvkprtehgektllanektlispgr--tehlg	1059
Qy	625	AKNOHEROLMETQCSIDINRIQSTRTADDCCVITVAANDGSDVYASVYEDTN-----	674
Db	1060	aktcsanekektlpsprakteh--gettlspndkltssaesctehndraasavilpapaerl	1118
Qy	675	--SQOKSLASQSTQKELOGLALATTOESPHPQNFQSTQEOQTHLRMEVNTIASSPLFS	732
Db	1119	khakrttlahckmq-----vteksstehpkektsttekt--trpektllysektlct	1169
Qy	733	HHDDQVLAEPTEHMGRRKDAKLTMEQFKATTRN--SPATTCGA--QFRP-----GLOAV	783
Db	1170	kykhnprpkeptelnlg---tlttelikarpvksenpekttaavtklpsvkvtygdksl	1226
Qy	784	DLTSTHV 790	
Db	1227	tttshl 1233	
RESULT	6		
ABB20314			
ID	ABB20314	standard; Protein; 1325 AA.	
XX	AC	ABB20314;	
XX	DT	23-JAN-2002 (first entry)	
DE	XX	Protein #2313 encoded by probe for measuring heart cell gene expression	
XX	KM	Human; gene expression; heart; microarray; vascular system;	
XX	KW	cardiovascular disease; hypertension; cardiac arrhythmia;	
XX	OS	congenital heart disease.	
XX	PM	Homo sapiens.	
XX	PN	MO200157274-A2.	
PD	XX	09-AUG-2001.	
XX	PF	30-JAN-2001; 2001MO-US00666.	
XX	PR	04-FEB-2000; 2000US-0180312.	
XX	PR	26-MAY-2000; 2000US-0207456.	
XX	PR	30-JUN-2000; 2000US-0608408.	
XX	PR	03-AUG-2000; 2000US-0632366.	
XX	PR	21-SEP-2000; 2000US-0234687.	
XX	PR	27-SEP-2000; 2000US-0236359.	
XX	PR	04-OCT-2000; 2000GB-0024263.	
XX	PA	(MOLE-) MOLECULAR DYNAMICS INC.	

XX	PI	Penn SC,	Hanzel DK,	Chen W,	Rank DR;
XX	DR	WPI: 2001-488899/53.			
XX	PT	Single exon nucleic acid probes for analyzing gene expression in human hearts -			
XX	PS	Claim 15; SEQ ID No 22084; 530pp; English.			
CC	CC	The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see CC AB21355-ABA41305). The present sequence is a protein encoded by one such CC probe. The probes may be used for predicting, measuring and displaying CC gene expression in samples derived from the human heart via microarrays. CC By measuring gene expression, the probes are useful for predicting, CC diagnosing, grading, staging, monitoring and prognosing diseases of the CC human heart and vascular system e.g. cardiovascular disease. CC hypertension, cardiac arrhythmias and congenital heart disease. CC Note: The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pcl_sequences.			
CC	CC	Sequence	1325 AA;		
XX	XX	Query Match	4.0%; Score 220; DB 22; Length 1325;		
XX	XX	Best Local Similarity	19.4%; Pred. No. 9.7e-08;		
XX	XX	Matches 153; Conservative	125; Mismatches 311; Indels 198; Gaps 38;		
QY	DB	113	DKLTKSDNGTARTLPAKONGTSDGCTIFVSTFVPASVSGKVSPTSOSGKNADRS	172	
DB	536	dtlplanekt	ttplslaeptengq-----tltplanektssaseptehertplanen	586	
QY	173	TLPRKSVQEGNDKCNAPSGKNGAALANDSPMKDLOGPANTDVANSEDNTSDVGCAL	232		
DB	587	tlp-spaetlent	-----ertanentltpsp---agptenemcan--ektllfpaep	632	
QY	233	PEVPOITWIEVNGADQPS-----TPKSEVVLKRNEDENGKTEETLVAEGCNLTKP-	286		
DB	633	tenretanekt	tspsaeptengrtqfrfanekttspsaeptehertplane-ntllapa	691	
QY	287	NPMGKE-----RDQVAGCNLTKP-----KPVSGCKEQJCN	320		
DB	692	epentretanekt	tlpfpaeptenertane-ntpsrqptengdrtpplanektjsla	750	
QY	321	EPCEEVLLKRSKSKRKTKLMKQOHSKRRKTAQADVDALCRKRPKVLLSPITIA	380		
DB	751	eplentg-krtfpfanekt	tsasaeptehera-----tplane	785	
QY	381	NOVEDSRSDVEVHRENA-----ADPCEDRS-----TIPVMEVSMDFVSNH--TV	424		
DB	786	nttspsaepentretanekt	qfpfaeptenestanektlfpae-----ptenrewta	840	
QY	425	GEBDLKSSKNKTKRKYSDDVDDGSSLMWNLGKKKRTGSHHTVAHPACN-----LSNKK	479		
DB	841	neuttlspaepteh-----eemtpfan-----ektlls-----peeptengertpfleak	885		
QY	480	VYPTAS--TOHDDENTENGLODTNMHKTTCVQHSVISIQRQ-----SSKGTAGLASK	530		
DB	886	tlpsaeptehertplane	l-tlpsaeptehertplanekttspsakptehgett-vte	943	
QY	531	GKTHSAASKTKYGG-----ESTRNGONIHVLSAEDOCQMETENSVLSHSAK	575		
DB	944	dtlpsaepentgerplanent	ttspstestehertanekt	tlppaepentgerptasn	1003
QY	576	-----VSPAEDHDIQMSDLHEQSILPKKKK---OKLEVTREKOTMIDIDIMDIVELL	624		
DB	1004	ektlpsakpteha-emptsanent	tlpsvkvptehgektllanektllspsgr--tehg	1059	
QY	625	AKNQHROLMTEHDCDINRIQSTKTAADDQCVYAAKQSDSASSVFDTN-----	674		
DB	1060	aktisanekt	tlpsakpteh-gerltspndkltssaseetehdratatsanvltppaepi	1118	

QY	675	--SOOKSLASGOSTOKELOGLHALTQOESPHQNPPOSTOEOCHLMEEWYTTAASSPLFS	732
Db	1119	khakrttlahekmtq-----vteksstehekttsttekt--rtpekrillysekiot	1169
QY	733	HHDDVYLAEPFHEHGRDAKKLMEQKATRN--SPAATGCA--QFPR-----GIAOV	783
Db	1170	kqknpryepkeptenlgn---tlltletikapykstenpektaavcklikpsvkvvtgdxsl	1226
QY	784	DLTSTHV 790	
Db	1227	tttsehl 1233	
RESULT	7		
ID	AAM55707		
XX	AAM55707	standard; Protein: 1325 AA.	
XX	AC		
XX	AM55707;		
DT	05-NOV-2001	(first entry)	
DE			
XX		Human brain expressed single exon probe encoded protein SEQ ID NO: 27812.	
KW		Human: brain expressed exon; gene expression analysis; probe:	
KW		microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;	
KW		epilepsy; cancer.	
XX			
OS	Homo sapiens.		
PN	MO200157275-A2.		
PD			
XX	09-AUG-2001.		
PF	30-JAN-2001; 2001WO-US00667.		
XX			
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX			
PA	(MOE-) MOLECULAR DYNAMICS INC.		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
DR	WPI; 2001-483446/52.		
XX			
PT	Single exon nucleic acid probes for analyzing gene expression in human		
PT	brains -		
PS			
XX	Example 4; SEQ ID NO: 27812; 650pp + Sequence Listing; English.		
XX			
CC	The present invention provides a number of single exon nucleic acid		
CC	probes which are derived from genomic sequences expressed in the human		
CC	brain. They can be used to measure gene expression in brain cell samples,		
CC	which may enable the diagnosis and improved treatment of nervous system		
CC	diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,		
CC	epilepsy and cancers. The present sequence is a protein encoded by one of		
CC	the probes of the invention.		
XX			
XX			
XX	Sequence 1325 AA;		
QY	Query Match	4.0%; Score 220; DB 22; Length 1325;	
	Best Local Similarity	19.4%; Pred. No. 9.7e-08;	
	Matches 153; Conservative 125; Mismatches 311; IndeIs 198; Gaps 38;		
QY	113	DKLRTSDNGTAPRTLPKANGTSDGCTITFVNSTFVPVSVGSQKYSPTQSSQGNADRS	172
Db	536	drtlanaktpslaeptengq-----rtflanehtktsasaeptehetrtplanen	586

QY	173	TLPSVSGDNDKSNAPSGKNGAEAENTDTPMKDLOGPQNTDVAAANSNDTSTVDAL	232
Db	587	tlp-spaerlen-----erlanentlpsr---agplehremlan---ektllfpaep	632
QY	223	PEVQIIMHIEVNCAQDPPS-----PPKISEVYLKKNEDNCKTEETUYAEQCNLTXP-	286
Db	633	tenreretanekttspaerplengqrtrpfaneekttsspaerlehgertprlane-nllispa	691
QY	287	NPMGKE-----RDQVAEOCMLTKDP-----RPVSGOKCEQION	320
Db	692	eptenretanekttprpaerplenretane-nltprsqrglengdrtrpaneekttprla	750
QY	321	EPCEEVYLKSSSKSKRRTKDKLKKQOHSKKRTPAQADVSDAKLCRRKPKVALLSETINA	380
Db	751	epteng-krtrpfaneekttssaaerlehaer-----trpane	785
QY	381	NOVEDSRSDVEHRENA-----ADPCEDRS-----TIPMEVSMIDPVSNH--TV	424
Db	786	nttspaerplenretanekttqfpaerplenrestanekttprpae-----ptenrewa	840
QY	425	GEDGKSSKNKTKRKYSDDVYDDGSSLNMWLNKGRKRTGSVHHTVHAHPAGN-----LSNKK	479
Db	841	neentlspaerlen-----eemtrpan-----ektlls-----peerlengertprfneek	885
QY	480	VTPPAS--TQHDENDPTENGLDITNMHKTVDYCOVYSISTORC-----SSGKTAGLSK	530
Db	886	tlpsaaerlehgertprlane-l-tlpsaaerlehgertanekttspaxplehgelt-vne	943
QY	531	GKTHSAASTKTKG-----ESTRNGONIHVLSAEDQCOMETENSYSLSAK	575
Db	944	dltpsaaerlengertprlanentltspstlehgertanekttspaerlehgertpsan	1003
QY	576	-----VSPAHDIQIMSDLHEQSLPKKKK---OKLEVTREKOTMIDIDIMDIVELL	624
Db	1004	ektlpsaerplehe-emptpanentltprpkrplehgertllanekttlsppegr---teh	1059
QY	625	AKNOHERQMLMETTSCDSINRIQSTTADDDCVYAAKDSDVASSYFDN-----	674
Db	1060	akttsanektlpsakpleh-gerltspndkltssaaestehdratsanvltppaepi	1118
QY	675	--SOOKSLASSTOKELQGHIALTTPOSPAPONFQSTOQOHTLRLMEEWVTAASSPLFS	732
Db	1119	khakttltahekmtq-----vtekstehpektsttekt--trpeprtllysektllct	1165
QY	733	HHDDOYIAEAFTHEWGRKDAKLLTWEQFKATTBN--SPAATCGA--QFRP-----GIGAV	783
Db	1170	kgkntprpexpletnlgn--tlltletlkapvstenpekteaavctklkpsvctvgdksl	1226
QY	784	DLTSTHV	790
Db	1227	tltsahl	1233
RESULT 8			
AAM68085			
ID	AAM68085 standard; Protein; 1325 AA.		
XX	AAM68085;		
XX	06-NOV-2001 (first entry)		
DE	Human bone marrow expressed probe encoded protein SEQ ID NO: 28391.		
XX	Human; bone marrow expressed exon; gene expression analysis; probe;		
KM	microarray; cancer; leukaemia; lymphoma; myeloma.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	WO200157276-A2.		
PN	09-AUG-2001.		
XX			

PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR.
 PI WPI; 2001-488900/53.
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 28391; 658bp + Sequence Listing; English.
 PS
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 CC
 SQ Sequence 1325 AA;

Query Match 4.0%; Score 220; DB 22; Length 1325;
 Best Local Similarity 19.4%; Pred. No. 9,7e-08;
 Matches 153; Conservative 125; Mismatches 311; Indels 198; Gaps 38;

OY 133 DKLKTSNDGTPRLTLPAPNGSDGCSITFRSTFPVAVSGSKSPSTQSSQGNADRS 172
 DB 536 dtrplanektspslaepengq-----rtplanektssaepeheerplanen 586
 OY 173 TLPKSVQBGNDKCNAPSGKNGAEANTDSPKLDGPAONTDYAANYSEDTSYDVAL 232
 DB 587 ttp-spaepren-----ertanenttppsp---agptenrean---ektllfpaep 632
 OY 233 PVPVQITTHIEVNGADQPPS-----TPKLSEVVLKRNDENGKTEETLVAECNLTKDP- 286
 DB 633 tenertlanektsspaepengqtrtplanektsspaepheertriplane-nttlspa 691
 OY 287 NPMGSKKE-----RDOVAEQCNLTKDP-----KPVSGQKCEQICN 320
 DB 692 eptenreretanekttrpfaepentertane-nttspaqptengdrtriplanektsp1a 750
 OY 321 EECCEVYLKRSSKSKRKTDKLMKQOHSKKRTAQAADVSDAKLCRRKKRYVALLSEIINA 380
 DB 751 epteng--ktrtplanektssaepehaer-----pplane 785
 OY 381 NOVEDSRSDVEVRENA-----ADPCEDRS-----TIPVMEVSMIDIPVSNH--TV 424
 DB 786 nttspsaepentertanekttrqfaepentertanekttrpfpe-----ptenrexta 840
 OY 425 GEDGLKSSKNKTKRKXSDVDDGSSLMNWLNGKKRGTGSHHTVAHPAGN-----LSNKK 479
 DB 841 neentlspapeh-----eemtrplan-----ektlls-----paepengertpfneek 885
 OY 480 VVPITAS--TQHDDENDTEGLTDMNKHKTVCQVHSEISYORC-----SSGKTAGAGLSK 530
 DB 886 ttpssaepheertriplane1-ttspaepheerlanektspakptehgettl-vne 943
 OY 531 GKTHSASTRKYG------STFNGQNIHVLASDCCOMETENSVLSSHAK 575
 DB 944 dttspsaepentertanenttststehertanekttrpfaepentertan 1003
 OY 576 -----VSPAEDHDIQMSDLHEQSLPPKKKK---QKLEVTREKQTMIDIDIMDIVELL 624

DB 1004 ektlspakpteh-eemtsanenttspvktpehgentllanektllspegr---tehng 1059
 OY 625 AKNOHERQIMTETDCSDINRIOSKTTADDCCYVAAKGDSDYASSVFDTN----- 674
 DB 1060 aktlsanektllspakpteh-gerltspndklttsaaestehtdrataaanvltppapepi 1118
 OY 675 --SQOKSLASQSTQKFLQGLHALTTOESPONFQSTOPOQHLMEEMVTTAAASPFS 732
 DB 1119 kktktllahemktq-----vtektsthektststekt--trpektllysektlct 1169
 OY 733 HHDDOYIAEAPTEHMGKRDAAKLTWEQKATRN--SPAATGCA--QRRP-----GIAQAV 783
 DB 1170 kgkntpvtpektenlgn-----ttltetlkapykstenpektlaavtklksvkvtdksl 1226
 OY 784 DLSTHV 790
 DB 1227 tltssnl 1233

RESULT 9

ID AAM03645 standard; Protein; 1325 AA.

AC AAM03645;

DT 09-OCT-2001 (first entry)

DE Peptide #2327 encoded by probe for measuring breast gene expression.

XX Probe; human; breast disease; breast cancer; development disorder;

KW Inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

PN W0200157270-A2.

PD 09-AUG-2001.

PF 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-476286/51.

Novel single exon nucleic acid probe used to measuring gene expression

in a human breast -

Claim 27; SEQ ID NO 12385; 322pp; English.

The present invention relates to novel single exon nucleic acid probes (see AAI00010-AAI10067). The present sequence is a peptide encoded by one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridizes at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer, disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.


```

Db      358  svssnadinnkkkttssnnngnnsaanknadiksnadlsatsnnaind-----413
      587  SDHEOSLPFKKKKKO-----KLEVTREKQTMIDDIPMDIVEL-----AKNQ 628
      414  -dheanssekptkadtfaarlatavagenelsdeefvyesaansknllffpsassqqq 472
      629  HEROLMETDSCSDINRIQKTTA---DDDCVIYAAKDGSPYAS-----SVFPTNSQ 676
      473  qqqqppqgqqqghgtsktsapllnnkkllsrlnksrhstgaallmhtltsltpn 532
      677  QKSLASSTCKELQGH-----LALTQESPHQNFQSTQEQHNLHMEKWTIAASPLF 721
      533  lsnvmqnnmlmsghhldelsikqepn-----qlqqqppmdvqsvdstsdp-- 565
      732  SHHDQYIAEPTHEWGRKAKRLTWEQFRATYRNSPAATC-----GAQFRGIAVDL 785
      586  ---dsnvakspkrrslvlskvsphllsttsngntlscpnvatnsqlepn---ndl 639
      786  TSTHVMGSS-----SNVASRQ-----PVAPLDRA-----ER 813
      640  stkslsnslrhnssanrnsgdnkrlptvtvsklfdsnpnqaplrtysgvpdhnled 699
      814  AVNOVHARNPSTIATWEASKICDRNAGOVLYPKESMPATHLRMDPSTLASFPNYG 873
      700  yleqph--nyptnqnsyvkdeefynsn-----nkfp--hglntfygdannyleeng 746
      874  TSS-----RNMESQLHNSOYAH-----891
      747  dsnnvrpqhnlqhefipedsenddlsmfyhknkndletkpllsdygededvdyd 806
      892  ---NOYGSTSTSGSNLNGKIRPL---TFEDL-----SRHQLHDHRLRPHPR 934
      807  rpnatfysygsaasnclheplhgrmpsrnndyddfmgvntnnglne-yuprlmkrg 865
      935  VGVLSLQKEIAMSSENGC-----TOSGYKLGVSTGITSHQW-NRKEHEBALNSGMSA 988
      866  grhl-srtmslmngslmngnddvthslmnddlvgysphnfysrkspfvkx----- 918
      989  KWNALQUGSVSS-----ADFLSARNSIAQSTMRGKGVHHLDRVKR-----DIC 1035
      919  -nelylafvisslmtgfljgfllatnkelqdv-----vvvmhdvissdsdelldlt 970
      1036  ITNKNPADFT 1045
      971  vsafnpgfts 980

```

RESULT 11

AAW20056
ID AAW20056 standard; Protein: 1528 AA.

```

XX AC AAW20056:
XX DT 10-SEP-1997 (first entry)
XX DE C. elegans UNC-53 protein variant 8A.
XX KW UNC-53; neuronal regeneration; revascularisation; wound healing;
XX KW neurodegenerative disease; Alzheimer's disease; Huntington's;
XX KW peripheral neuropathies; metastasis inhibition; cancer.
XX OS Caenorhabditis elegans.
XX PN WO9638555-A2.
XX PD 05-DEC-1996.
XX PF 31-MAY-1996; 96WO-EP02311.
XX PR 31-MAY-1995; 95GB-0010944.
XX PA (BOGA/) BOGAERT T.

```

```

PA (STR/) STRINGHAM E.
PA (VAND/) VANDEKERCKHOVE J.
PI Bogaert T, Stringham E, Vandekerckhove J;
DR WPI; 1997-034369/03.
DR N-PSDB; AAT71314.
PT Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
PT promote neuronal regeneration, revascularisation or wound healing.
XX Claim 20; Page 106-111; 278pp; English.
XX UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.
CC The UNC-53 proteins and nucleic acids are useful as medicaments to
CC promote neuronal regeneration, revascularisation or wound healing, or
CC for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
CC Huntington's disease) or acute traumatic injuries. Transgenic cells and
CC organisms transfected with UNC-53 cDNA can be used to determine whether
CC a substance is an inhibitor or enhancer of the regulation of cell shape
CC or motility or the direction of cell migration by screening for a
CC phenotypic change in the cell. Inhibitors can be used to alleviate the
CC spread of disease inducing cells or metastasis. Probes derived from the
CC cDNA sequences can be used to identify homologues of the C. elegans
CC unc-53 gene. The UNC-53 protein can be used to identify proteins which
CC are active in the signal transduction pathway that can be used as
CC mentioned above.
XX Sequence 1528 AA:
SQ

```

Query Match 3.8%; Score 208; DB 18; Length 1528;
Best Local Similarity 19.5%; Pred. No. 1.1e-06;
Matches 231; Conservative 174; Mismatches 427; Indels 350; Gaps 57;

```

QY 14 VGTNCM-----LARGTGAVAVLELTATPRRODAAEAGVDPAHQCHFSIRGVAL 67
DB 83 lgydcskltktdasgdlgavqlflfsltykq----- 115
QY 68 LQKKDPFCSLSRIFHQKCKDEKASSPSFSAKFRMDCSKLDKLTSDNGTAPRTL 127
DB 116 -----klrqlkdkqkqlleqpsimpavsklpsrvatsatatsnpsnfqms 166
QY 128 PAK-QNGTSDGCTTFVSTFVPASVGSQKVSPTOSSQGNADRLTPKSVQEGND--- 183
DB 167 tsrltpqsriskidskiglkpkslsglkpsssttsnntnsfr---psrsgnnny 223
QY 184 -----SKNAPSGKNGAENATDSPMKDLOGPQONTVVAANVSEDMTSVVGALPEVPI 238
DB 224 ststssakslssstyslsnlnlpsqlkpsrptqlvrva---ttklgs----- 273
QY 239 TWHIEVNGADQPPSTPKLSEVY---LKRNDENKGTETELVADQNLTKDP-----NPM 290
DB 274 -----sklaapkvastpklavsktlgkqepdngsggggmlklklfasknpsssnspq 329
QY 291 GKERDQVAEQCNLTQDKRPV-SGQKCEQICNEPCEEVYVLRSSSKS-----RKTDK 341
DB 330 lrkaavpqqqlsklaapyksglk-----pptsklgsatmskclcpkvsyrtkldap 382
QY 342 LMKRQGSKKRRTAADVSDAKLCRRKPKK---VRLSEIIMANOVDSRDEYHRENA 397
DB 383 llsqg-----dktrcksseesgygfnstpsstsgsls---mhsstsk 427
QY 398 DPCEDDRSTIPVMEVSMDI-----PVSNHTVGED-----GLKSSKNKTKRK 439
DB 428 stsdkspsddltlnaslvtairgplaavrpynlfnkpvveekrlavkykstakd 487
QY 440 KSDVVDGSSSLMNLNGKKRTGVSVHTVAHPACNLNKKVT--PTASTQHDDENDTENG 497
DB 488 pppavpprpt-----qrlgvvpspiman-----kklndrvlsekpeklsgms 532
QY 498 LDTNMHKTDV--COHVSEISTQRCSSKGTAG---LSKGTSHASTKYGGESTRNGON 551

```

```

Db 533 idt-----tdvprlpplksvlpkmtslrppdyvllkggkltspvks-fgyeqs----- 582
QY 552 IHLVSAEDOCOMETENSVLSH-SAKVSP-----AEHDIQIMSDLHEOSLPKRRKKOLEV 605
Db 583 -----sasedslvahasavtprpcktsqghsl-----ermgknktsessgy 624
QY 606 T-----REKOTMIDIPMDIVELAKNOH-----EROLMETEDCDINRIOSKTT 650
Db 625 tsdagvamckmreklkeydmdmr-----ragngypdnfcdssalsqslsdmlddist 679
QY 651 AD---DDCVVAAKDGSDVYASV-FDTNSOOK-----SLASQSTOKELQGHIALTTQESPH 702
Db 680 ddlsydmavtsakh-sdyshfvrhpsstsskprvpsrstsvdsrarae-----genvy 733
QY 703 PQNFQSTQEOQTHLRMEEMVTIAASSPLFSNH---DDQYIAEAPTHEWGRKDAKXLTWEQ 759
Db 734 klisqctsqrg-----aaatslqghslrppgyssyp--hls-vsadkdlmsm 780
QY 760 FKATTR--NSPATCGAOPRPGIOAVDLTSTHWGSSSNVYASROPVIAPLDRYAERAVNQ 817
Db 781 hqgstrpsaqkpsysqf-----hsld-rkchlfqfctshmaalls----- 824
QY 818 VHAANPSTIAITMEASLCDRNAG--QVVLTKRESMPATHLRAMPSTLASPPNKTGS 875
Db 825 ---rrypnsmkydssgysarsrgsstlgye-----tfqlhrlsdckspah-----s 871
QY 876 SRNOMESOLNSQYAHNQYGSTSTSYGNLNCKIRPLTFEDLSRHOLDLRHPL----- 929
Db 872 akemsqgl-----slaatlqys-lneky-----ehairdnardlecqknt 912
QY 930 ----RPHRVGVLSLQKEIAMSSENCSTQS-----GKVLGVSTGITSHQ 971
Db 913 vdslltkqgenyalfdlfeqklrklctqldtrnlkpeearlrfqdlahldismhlaasns 972
QY 972 MNKKE-----HEALNSGMFSAKMNAILOGSVSSSADFLSARSTIAQSW 1015
Db 973 ahneagagellirpslesvashrsmsssskskqeklslstlqk-----ksw 1022
QY 1016 TRGKGKVVHPLDRFVRDICTNNK--PADFTTISDNENEYMD 1055
Db 1023 trss-----lskftkk-----kknkydeaimpslssgsqcl 1054

```

RESULT 12
ID AAM20057 standard; protein; 1583 AA.
XX AAM20057;
XX
XX
XX 10-SEP-1997 (first entry)
XX
XX C. elegans UNC-53 protein variant 7A.
XX
XX UNC-53; neuronal regeneration; revascularisation; wound healing;
XX neurodegenerative disease; Alzheimer's disease; Huntington's;
XX peripheral neuropathies; metastasis inhibition; cancer.
XX
XX Caenorhabditis elegans.
XX
XX W09G638555-A2.
XX
XX 05-DEC-1996.
XX
XX 31-MAY-1996; 96NO-EP02311.
XX
XX 31-MAY-1995; 95GB-0010944.
XX
XX (BOGA/) BOGAERT T.
XX (STR/) STRINGHAM E.
XX (VAND/) VANDERCKHOVE J.
XX
XX Bogaert T, Stringham E, Vandekerckhove J;
XX

```

DR WPI: 1997-034369/03.
DR N-PSDB; AAT71315.
XX
PT Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
PT promote neuronal regeneration, revascularisation or wound healing.
XX
PS Claim 22; Page 111-116; 278pp; English.
XX
CC UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.
CC The UNC-53 proteins and nucleic acids are useful as medicaments to
CC promote neuronal regeneration, revascularisation or wound healing, or
CC for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
CC Huntington's disease) or acute traumatic injuries. Transgenic cells and
CC organisms transfected with UNC-53 cDNA can be used to determine whether
CC a substance is an inhibitor or enhancer of the regulation of cell shape
CC or motility or the direction of cell migration by screening for a
CC phenotypic change in the cell. Inhibitors can be used to alleviate the
CC spread of disease inducing cells or metastasis. Probes derived from the
CC cDNA sequences can be used to identify homologues of the C. elegans
CC unc-53 gene. The UNC-53 protein can be used to identify proteins which
CC are active in the signal transduction pathway that can be used as
CC mentioned above.
CC
XX
SQ Sequence 1583 AA;

```

Query Match 3.8%; Score 208; DB 18; Length 1583;
Best Local Similarity 19.5%; Pred. No 1,1e-06;
Matches 231; Conservative 174; Mismatches 427; Indels 350; Gaps 57;

```

QY 14 VGTNCM-----LARGTGAVAPVLELTATPRDAAAAGVDEPAHQCEHFSIRGYAL 67
Db 83 lqdcskltktddidsgnlqglvllfllstkyq----- 115
QY 68 LQKDPKFCSLRFHQKCKDEHKASSPSPVAKFRMDCSKLDLKTSDNTAARTL 127
Db 116 -----klrqlkdkqkleglqltslmpavaklspyratatasatprnfnfpm 166
QY 128 PAK-ONGSTDGCSITFRSPFVPSVGSOKVSPSTOSSOGKNADRTLPKSVQGNP-- 183
Db 167 tsrlqprstlksldsklqkpktsqlkppssstssntnstr---psarsgmnngv 223
QY 184 -----SKCNAPSGKNGAAEANTDSPMKDLQPAQNDVAAVNSBDNTSVYDGLPEYPOI 238
Db 224 stlstsakslssstyslslnlnrptqlqkpsrpqqlrvva---lttkigs----- 273
QY 239 TWHTIEVNGADOPSTPKLSEVY---LKRNEDENGKTEETLVAEOCNLTQDP-----NPM 290
Db 274 ---sklaapkvastpklasvktlgakqepdnsgggg99gmllklflfasknpssssnpgp 329
QY 291 GKERDOVAEOCNLTQDPKPV-SGOKCEQICNEPCEVYLKRSKSK-----RKTDK 341
Db 330 trkaavpqqqltsklaapyaglk-----pptsklgsatmskklctprkvsytkldap 382
QY 342 LMKQOQSKKRTQAQADVSDAKLCRRKPK-----VRLSEITIANQVEDSRDEYHRENA 397
Db 383 llsqg-----dktrcksaseesgygfnstprtsstsgsls--mhsatsk 427
QY 398 DPCEDDRSTIPVPMEVSMDT-----PVSNHTVGED-----GLKSSNKKTKRK 439
Db 428 sstsdckspssddltlnaslyvtairgplaavprnlnkpvcekrlavgyvxtakd 487
QY 440 YSDVYDDGSSLMWNLNKKKRTGSVHNHTVAPACNLNKKYV--PTASSTQDDNDENDENG 497
Db 488 ppaavprdt-----qrltgvpslmdh-----kltlnrvlsekepekllgms 532
QY 498 LDTNMHKTVDV--COHVEISITPORSCKKTAG-----LSKGTTHSAAGTKYGCSTFRNGON 551
Db 533 idt-----tdvprlpplksvlpkmtslrppdyvllkggkltspvks-fgyeqs----- 582
QY 552 IHLVSAEDOCOMETENSVLSH-SAKVSP-----AEHDIQIMSDLHEOSLPKRRKKOLEV 605
Db 583 -----sasedslvahasavtprpcktsqghsl-----ermgknktsessgy 624

```

```
QY 606 T-----REKQTMIDIDIMDIYELLAKNOH-----ERQIMTEIDCSINRIQSKTT 650
DB 625 tsdagvamcakmreklkayddmtr-----raqnydpdiefdsisslsjsidnneidist 679
QY 651 AD---DDCVIAAKNGSYASV--EDTNSOOK-----SLASOSTOKELGHALTLTQESPH 702
DB 660 dslsgvdmavtvaakn-sdysnfvrhptsssskprvpsrtrsvdrtrae-----qenvy 733
QY 703 POFPOSTEOQTHLMEMEAVTIAASSPLFSHH---DDQYIAEAPTEHMGKRDARKLTWEQ 759
DB 734 klsgcrtcsqrg-----aaatstfgqhsrlspysysp--hls-vsadkdtmsm 780
QY 760 FKATTR--NSPATCGAOPREICAVDLSTVWSSSNYSASROVIAPLDRYARAVNQ 817
DB 781 hsqtrsrpsqkpsysgdf---hsld-rkchldqfstehmaalisp----- 824
QY 818 VHAARPSTIATMEASKLCDRRNAG--QVLYPKRSMPTATHLRMMDPSTLASFPNGTS 875
DB 825 ---rtvpmmskydssgsysatsrsgsstglye-----tfqhrisdeksph-----s 871
QY 876 SRNQESQLNSQYAHNOYKSGSTSYSGNLNGKIPLTFEDLSRHQLDHLRPL----- 929
DB 872 aksengsqj-----slastltaggs-lneky-----ehairdmardlecyknt 912
QY 930 -----RPHRRVGLSLLOKETIANNSENGTQS-----GYLGVSTGITSHQ 971
DB 913 vslstlckqenygalldlfeqklrklqldrsnlkpeearlfrqdlahlrdisnhlaasns 972
QY 972 MNRKE-----HFEALNSGMSAKWNALOLGVSSSADFLSARNSTIASQSM 1015
DB 973 ahaneagagellrqpaleavashrsmsasssksgkqekialsfgkn-----kkaw 1022
QY 1016 TRGKGKMHVPLDRFVADICITNKN--PADFTTISNDNRYMD 1055
DB 1023 lrs-----lskftkk---kknkyaedahmpsisgsgtld 1054

RESULT 13
ABG03703
ID ABG03703 standard; Protein; 2070 AA.
XX
XX ABG03703;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #3694.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX
XX (HYSE-) HYSEO INC.
PA
XX
XX Drmanac RT, Liu C, Tang YF;
PI
XX
XX WPI; 2001-639362/73.
DR
XX
XX N-PSDB; AAS67890.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
1
```

```
XX
PS Claim 20; SEQ ID NO 34062; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2070 AA;
XX
XX
XX Query Match 3.7%; Score 206.5; DB 22; Length 2070;
XX Best Local Similarity 19.1%; Pred. No. 2.3e-06;
XX Matches 231; Conservative 188; Mismatches 509; Indels 279; Gaps 53;
XX
QY 1 MEIVA-----VQEGARVYGTNCLMARGGTGAVYELTATPRDAAEAGVDEPAHQOC 56
DB 920 velleiporvegdssdmtqtlgllkqgglltclldlakkknpgkslsdq-mhpr----- 973
QY 57 EHFSIRGVALLQKDDPFCSLRFH-----DQKKCDEH----- 91
DB 974 -dfs-----sepnklydssvnlhpkqelllmmnddrppqhsclpdevtngsl 1023
QY 92 --KASSPFSVA--KFRWDCSKCLDKLTSDNGTAPRTLPAKONGTSDGC-STFEVRSST 146
DB 1024 ngrqdsdpmtsgsrsfsvasmlpetred-----vtsnatntcdscftveln 1074
QY 147 EYPAVSGQKVPSTQSSGCKNADNSTLPKSYQEGNDKCNAPSGKNGAALENDTSPMKD 206
DB 1075 xysscsksxyfxp-----gepxegeswpqxykxssskpseaslllegdppfksgipkes 1128
QY 207 LOGPQNVYVANNVSEDNSTVDVGLP-----EVPQITWHIEVNGA-----DQPSSTPKL 256
DB 1129 gtcgae-----atpnefngsgeiatemerplekpscsjgiktsnslsldstsqpslsl 1183
QY 257 SEVVLKRNED-----ENGKTEETLVADQCNLTQDPNPMSEKENDQVAEQCNLTKD 306
DB 1184 svnnllngsishplascaglsptsegtvtvpatvnlvssasygsqpprlsdvlnl--- 1240
QY 307 PKPVSGQCEQICNEPCEEVVLKRSKSKRKTDKMLMKQOQSHKRRQAD--VSDAKL- 363
DB 1241 pknsxipmtstlpsnqilqepillkpsnesrk-----dsakrvagddlllsakrg 1289
QY 364 --CRRKPKRVRLSEITINNOVEDSRSDVHRENADPCEDDRSTIPVMEVSMIDPVSN 421
DB 1290 knqcpaprlleaslm---stprdtlsdgtqmmvsglppnsasv-----vpyan 1336
QY 422 HTVGEDGL-----KSSKNKTKRKYSDVVDGSSLLMWLNGKKRKGSGYHNTVAHPAGNIS 476
DB 1337 pahg-dgltrlfrpsnflvpralrqlrevqsgsp---vaeqqltasqhlqasaaacs 1392
QY 477 NKKVTPPQAST--QHDDENDTENGDLTMHKTQDVCOH-----VSEIST 516
DB 1393 srclspstlxstxssssssssssnmmkqgsxerjtylksimylmqralstslsp 1452
QY 517 QRCSSKGTAGLSKGKTHSAA-----TKYGESESTRN-----GONT 552
```

```

Db 1453 lmslrsgatgscdeekescscgphdlsavpnyngsmvgtdgrsktqgphhgmqgm 1512
Oy 553 HVLSEDDCQMTENSVLSSHAKVSPAEDHDIQMSDLHQSLEPKKKKKLEVTREKQTM 612
Db 1513 qghfsgsqtekcscenstsrnhnhpqn---lmgdlmqgdvsgsgsgvseeh--v 1567
Oy 613 IDIDPMDIVELLAKNOHEROLMTE-----TDCSDI-----NRROS-----KTTA 651
Db 1566 sglnpmq---rltsrglbeqmwsgpsivtrssdmctcphrpernvssysaeallgkts 1625
Oy 652 DDDCVIAKDCSDYAS-----SVEPTNSQOKSLASQSTQKELQCHLAVTQESPPHQN 706
Db 1626 nseqtmglsgsvsdqlemtsyldv--prnkslahmmgtrvd--htvasdrlsdqctf 1683
Oy 707 Q---STQEQGTHLMEHMTIAASSPLFSHDDQYIAEAPTEHMGKD---AKKLWEGP 760
Db 1684 kpsgasqgqsnfeygssrnelgnpvsrlrmgsqatrlsqntppridqrkrlsyppv 1743
Oy 761 KA--TTRNSPA--TCGAOPRPGIOAVDLTSTHWGSSSNVASROPYAPLDRYAERA 814
Db 1744 gslptgnglpsidsentcstqsfmgallaphl--sdqyisqslsehg-----ntqcg 1795
Oy 815 VQVHARNPSTIATMEASKLCDRNAGQVLYPKESMPATHLLRMMDSTLASPPNYGT 874
Db 1796 psaleynopclhenvhlrtseesq-----rescdms--lgaIntnslnhpfs 1846
Oy 875 SSRNMQESOLHN---SQYAHNOYKGGSTSYGSNLNGKIPLPFEDLSRIQDLHRLP 931
Db 1847 ssagdlgrntspnsvqsknmpmlteshatkghmp--pvt-----tmhgvvarpalp 1898
Oy 932 HRRV---GVLSLLOKELANNSENCQSGYKLGVSITGITHQNRKHEFALNSGMS 987
Db 1899 hpsvshngdggpavraq-----nsyvpqrsrhlpgdsgsklrqpernrsngqrslvd 1954
Oy 988 AKMNAQLQSSVSSADFLSARNSIAOSWTGKCGKWHVPLDRFVRODICITNKNPAFTFI 1047
Db 1955 plplphpl---stgsmllgrq---qpatckrgslv-----rfm-----pdsqpy 1993
Oy 1048 SUNDNEYM 1054
Db 1994 pndnsdm 2000

RESULT 14
ABB68397 standard: Protein; 2768 AA.
XX
AC ABB68397;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31983.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NT.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
WI; 2001-656860/75.

```

```

DR N-PSDB; ABL12500.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABR72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2768 AA;

```

```

Query Match 3.7%; Score 204; DB 22; Length 2768;
Best Local Similarity 19.7%; Pred. No. 5,6e-06;
Matches 181; Conservative 118; Mismatches 345; Indels 274; Gaps 40;

```

```

Oy 29 APVLELTATPPRODAALAEAGVD-----EPAHQCHEFTIRGVALL 68
Db 1550 apavltatpsskddgqkpvveekpledqkpledestdstpslaselepepdattlap 1609
Oy 69 OKKDPKFCSLRIFHOKKODEHKASSSPSVAKFRMDCSKLDKLTGNTONTAPRTU- 127
Db 1610 skeepsepsgapktde---paepsctapes-----deekpepeevpttva 1653
Oy 128 PAKONQTSQSGSITFVSTFVPASVGSOKVSPSTOSSOGKNADRSITLPKSVQBGNSKCN 187
Db 1654 pagex-----ipcsiltpeedprataapvakpdedv---eketstelpid 1695
Oy 188 APSGKNGAALAEANTDS-----PMKDLQCPAO--NYDVANVSEDNSTVDV-----GA 221
Db 1696 apaseeedsensldqipsevpekkpeltqptpeegdlvgataapltasdevpqrjipev 1755
Oy 232 LPEVPQITWHIEVNGADQSPSTPKL-----SEVLAKRNDENVGKTEETLVAE 278
Db 1756 laelpgstetgikqgdeltaapsldrkepyteldeeaactvapisekdekpreekpve 1815
Oy 279 QCNLTGDPNPMGSKERDOVAEOCNLTGDPKPVSGOKCEQICNPECEVYLKRSKSKRKT 338
Db 1816 qkptgeeps---eeekekplegdvsteg--pvtaseagstassev-----kpst 1863
Oy 339 DKRLMKKQOHSK-KRTAODAVS-----DAKLCHR--KPKKVRLLSEITIANOVEDSRSD 389
Db 1864 egveaekpedkqpsstaqavetlpeistelpaqdgdqkpsaapvdsdeltsdekkip 1923
Oy 390 EVHRENAADP-----CEDDRSTIPVMEVSM--DIVSNHTVGEDELKSKSKNKTTR 438
Db 1924 svsgveegpevltaspqaaeedelktpaesepstckkpetekyqkpedetka--detpe 1981
Oy 439 KYSDVVDDGSSLLNMLNGKKKRTGGSYHH-----VVAHPAGNLSNKKKYPTASTOHD----- 489
Db 1982 svtqvsvdaltastaprag-----gdlekedqataape-----eeekrltapiapaelqps 2033
Oy 490 -----DENDTENGLDITNMHKTQVCHVSEISTORCSKSKGTAG--LSKGTGHSAASTRYGG 543
Db 2034 ekepvdegevesgckaipaesd--gpidelap-----atsgpldeast--aapltk--- 2080
Oy 544 ESTRNQONIHVLSAEDQCQMETENSVLSSHAKVSPAEDHDIQ-----MSDLHEQSLP 595
Db 2081 -----eestlvaa--aspavhdelldvcttqpvadekevaap 2117
Oy 596 KKKKOKLEVTRKQTMIDIPMDIVELLAKNOHEROLMETDCSDINRIQSKTTAADDC 655

```

Db 2118 gdelktidsvdsptlaqdd-----ekgdkteapayptl--vssptladsaa 2161

Qy 656 -----VIYAKDGSIDYASVFDITNSQOKSLASOSTOKELGCHLALTQE----- 699

Db 2162 dssrltpevpssvldtkpmddlmsqtatphtad--gaaststedeqavtvsbpqdaekt 2220

Qy 700 --SPHPONFSTQOQOTFLMEEMVTIAASPLFSHDDQYIAAPTEHMGKDAKLLTW 757

Db 2221 pvsrpapdscktpssseapqdaelip--atatpl-----dd----- 2253

Qy 758 EGFKATRNSPAATCGAOFKPGIOAVDLTSTHMGSSSNAROPVIAPLDRVAERAVNQ 817

Db 2254 -----nkpractvaptddgyvat-----aapldedkigttaaptl----- 2289

Qy 818 VHARNPSTTATMEASKL 835

Db 2290 ---ekipstaapldedki 2304

RESULT 15

ABB60291

ID ABB60291 standard; Protein; 2951 AA.

XX ABB60291;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 7665.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO2001/71042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI: 2001-656860/75.

XX N-PSDB: ABL04394.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX

PS Disclosure; SEQ ID NO 7665; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABLI6176-ABLI0511), expressed DNA

XX (ABB5737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pcl_sequences.

XX

XX Sequence 2951 AA;

Query Match 3.78; Score 204; DB 22; Length 2951;

Best Local Similarity 20.1%; Pred. No. 6.2e-06;

Matches 213; Conservative 161; Mismatches 463; Indels 222; Gaps 48;

Qy 78 LSRIFHDQCKDEHKAASSPFSVAKFRWDCS-----KCL-----DK-LKTSNGT 122

Db 1838 vsrltpevpssvldtkpmddlmsqtatphtad--gaaststedeqavtvsbpqdaekt 1892

Qy 123 APRILPAK-----ONGTSDGCSITFEVNST--FVPAVSQOKVSTQSSQOK- 167

Db 1893 rptslprtelergplaeideellrddctslvsqkvkhfift---aeklapmpkspri 1949

Qy 168 --NADRSTLPKRSVEGNSDKCNAPSGKNGAAEANTDSFMDLQBPQAQ--NYDVAANYSE 222

Db 1950 vanterhlsrsgsepereldeeep-----eldrtdvdeddqtsgtleeteeltqvtk 2002

Qy 223 DNTSVDVALPBEVQIITMHEIVNGADOPRSTRPKLSEVVLKRNEDENGTEETLVAE-OCN 281

Db 2003 ket-----lkefxyqktelrettrdskaepklyq---kxspqtkvveesaarvpyqak 2052

Qy 282 LTKDPNPMGSKERDQVABQCNLTKDPKPSGQKCEQICNEPCEFEVVLKRSK-SKRYTDK 340

Db 2053 vsqkvsqweprkkqprekvteqkeplepkqplskvkdpep--ekvnrkrepkvqkesqt 2110

Qy 341 KLKKKQHSKRRRTAODVSDAKLCRRKPKVRLSEITINANQVEDSNDSEVHRENAADPC 400

Db 2111 klikepervtlktprq-----keprkepl-----rqsedepefspeefddepl 2153

Qy 401 EDDRS-TIPVPMENYMDIPVSNHTYGE--DGLKSSKNKTKRKYSVDVDDGSSLMNMLNGK 457

Db 2154 pmktnthtaemkrqkdlmnpvsvfgtrtperksastprpklngtrgrpsncilllee 2213

Qy 458 KKRFGSVHHTVAHPAGNLNKKVTPRTASTQHDENDENGIDTMMHK-----T 505

Db 2214 krsyngvtnvskp---gtrtktpsanspqspptktsiskmeqsgsqsvwvqdv 2269

Qy 506 DY-----COHVEISITQRCSSKGTAGLSKGTHS---AASRYGGESTRNQNIHYL 555

Db 2270 dveavvgpaprpshisekpgqkspstsrslsrpsrpsksiltly---tlntgriva 2325

Qy 556 SAEDOCOMETENSVLSHSAKVPSEHDIQMSDLHQEGLPKKKKOKLEVTR---EKOT 611

Db 2326 srinvfevhetvhdseptgtrrpsymd-----htkslehlrrdslelnksystrss 2378

Qy 612 MIDDIPMDIVELAKNQHERQMLTETDCSDINRIQSKTTADD--DCVIYAKDGSY--- 666

Db 2379 meddsprve-----prnpssvkvf-----dyrvkssrgradeprktslkygdedsdlele 2427

Qy 667 ASSVFDITNSQOK---SLASOSTOKELGCHLALTQESHPHNPFOSTOQOHLIMEEAVT 723

Db 2428 leelfldlqrllelvasyemrrrlraqmrlir-----kmlnagttlittlittstt 2481

Qy 724 IAASSPLFSHDDQYIAAPTEHMGKDAKLLTWEQFKATTNRNSPAATCGAOFKPGIOAV 783

Db 2482 pykessplrkirrdqspagaa-----evtkkerytlltsrrqgqgrvqv 2524

Qy 784 DLSTHMGSSSNATASROPIAPLDRVAERAVNOVHARNP-----STATMEASK 834

Db 2525 dslrtpapqktspsng--kprkpreresapqkr---flsppqyspqrddttlktkvt 2579

Qy 835 LCDBRRNQGVVLYLKESPA-----THLLR--MMDPSLASPNNGTSSRRNQMESQL 884

Db 2580 tstrtga-----pkrpaggprlwardskvlgahatvpqngstprkqssstss-- 2629

Qy 885 HNSQYAHNQYKSTSGSNMLNGKIRPLTFPDLRSHQLHDRLPRPRRGV-----LGS 940

Db 2620 --gklttrmtssstlssstltnrknqreedsltsy-----gvgrtdengrlflgira 2682

Qy 941 LLOREIANWSENCGTQ---SGYKLG---VSTGITSQNRKREHEALNSGFSAKWNLQ 994

Db 2683 lkkkatpraeepcetkqvgtvyvleeqfysdnkappriherkellysnaadelaaklqglq 2742

Qy 995 LGSVSSADPLSARNSTIAQSWTRGKGKWWPLD-REVRQ 1032

Db 2743 -deddspplldar--vvrefkvesqslpedayvtr 2778

Search completed: September 16, 2002, 22:23:16
Job time: 8005 sec

THIS PAGE BLANK (USPTO)

Db 1474 -----ESKSRRSVSQPKETSAETTVASTOETVDNSVSTPKPR 1514
QY 705 NFOSTQEOHTLMEKAVTTAASSPLFSHHDDOYIAEAPTEHNGRKAOKKLTWEQKATY 764
Db 1515 SRRTRRSVQTN-----SYEPV-----ELPTEN----- 1536
QY 765 RNSPATCGAOPFRGIAVDLTSTHVMSSSNYSARQPVYAPLDRAVRAVNOYHANFP 824
Db 1537 -----AENAVNO-----SGNNVANSOP-----ALRNLTSKNTN 1565
QY 825 STATMEASKLCORRAGVYLYPKESMPATHLRMDPSTLASFPYGT-----SSRNM 880
Db 1566 AISISNAAKAOFVALNGKAV-----SOHISQL-----EMNEGQVNWISMTSM 1610
QY 881 ESOLHNSQYAHNOYKGSTSYSGSNLNGKIPLTPEDELSRQHLHAPLRHPVGVYGS 940
Db 1611 -----NNKSYSEQYRRSSKSTOFOLMDOTIS-----NNVQLGCVFTYVRSNNFDRASS 1661
QY 941 LLOKEINMSENGCTOSGKLGVSCTGITSJHOMNREH-----FEALNSG 984
Db 1662 KNTLAQVNFYSKYADNHHWLTGLDYGKFOSNLQTNNAKFAHHTAQIGLTAGKAFNLG 1721
QY 985 MESAKMNAALDGSVSSADFLSARNSIAQSWTRGKGMVPLD--RFRODICIT 1037
Db 1722 NFAVKPTVGVRYSYLSNADPALADRI-----KVPDISVKTAFQVLDLSYT 1767

RESULT 3
US-08-425-061-20
; Sequence 20, Application US/08425061
; Patent No. 5622829
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Loti
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,061
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-425-061-20
Query Match 3.38; Score 182; DB 1; Length 900;
Best Local Similarity 18.4%; Pred. No. 1.2e-05;
Matches 200; Conservative 142; Mismatches 363; Indels 382; Gaps 49;
QY 48 VDEPAHOCEHFSIRGVALL--OKKDPKCSLSRIHDOKKOCEHKAASSPFS--VAKF 103
Db 31 IKPEVSTKCDHIFCKFCMLKLNCKGSPQCLKNDITKSLQE-----STRESQVLEL 86
QY 104 RRMDCSKLD----- 113
Db 87 LKICAFOLDGLEYANSYNPAKKNENSPHELKDEVISIOSMGYRNRAKRLJOSEPNPS 146
QY 114 -----KIKTSDNGTADPTLPKONGT-----SDGCSITFYRSTFVPAVSQK 156
Db 147 LOETSLVQLSNLGTV--RLTKTKORIOPKTSVYIELGSDSSEDTVKKATY--CSVDQE 203
QY 157 VSPSTOSSGKNADRLTPKSVQEGNSKCAPSGKNGAE--AMTDSFMKDLGPAONYD 215
Db 204 LLOIT-----PGTDELSLDSAKKAACESETDVTNTEHHOPSN-- 244
QY 216 VAAVSEDNSTVDYCALPEVPOITWHIEVNGADQPPSTPKLSEVVLKRNEDENGKTEETL 275
Db 245 -----DLNTT-----EKRAERHP-----EKYGSSVSNL 269
QY 276 VAEQCN-----LTKDPNPMGKERDOVAEOCNLTQDKPVSGK----- 314
Db 270 HVEPCGNTTHASSLOHENSLLTKD-----RANVKEAFECNKSOKPGLARSQHNWAG 323
QY 315 CEOJCNEPCEEVVLKRSKSKRKTD--KKLMKQOHSKKRTAQADVSDAKLCRRPK-- 370
Db 324 SKETCND-----RKTPSTKKVYDLNADPLCEKRMKKOLP-----CSENPROT 367
QY 371 -----VLLSEIINANQVEDSVHRENAADCEDDRSTIPVMEYSMDI-----PVS 420
Db 368 EDVPMITLNSIQKAVNE--WFSRDEL--LGSDSHGOSESNAKVADVLDMEDVEYS 423
QY 421 NHTYGEGLASSKN-----KTKRKYSDVYDDGSSLMNMLNKKRRTGVSVHTVAH----- 470
Db 424 GSSEKIDLLASDPHEALICKSERVSKSV--SNIEKIFPKTKKKSLSPLNLSHVNTL 481
QY 471 -----PAGNLSNKKVTPASTQHDDENDTENGDLTNMKTQVOC--OHSEI 514
Db 482 ITGAFVTEPQIIOERPLNKLKRRRPTSGLHPD-----FIKADLANOKPPEM 531
QY 515 ---STORCSSKKTAGLSKGTHTSAASTKYGSESTRNCONITHVLSAEPOCMETENSYLS 571
Db 532 INQGTNQTGONGYVMTI-----TNGHEHKTGDSIQNKKNPIES-----LEKESAFT 582
QY 572 HSAKVSPAENDIQMSDLHEOSLPRKKKKOKLEVETREKQTMIDIDPMDIVELAKKNOHER 631
Db 583 KAEPISISSNMELELNHNSKAPKKNRLRRKSSTRH-----IHALLVYVRSNLS 633
QY 632 QLMTE---TQOSDINRIQSK-----TTADDQCVIAKQSDVASSVFDNISOQKSLAS 682
Db 634 PNCTELQIDSCSSSEEEKKRYNOMPVHSHRNLDLMEGKEPATGAKKSNKPNEDTSKRHD 693
QY 683 QSTOKELQGLHALTQOESHPONF---OSTQEOHTL-----RMEVYTIASSPLF 731
Db 694 SDTPELK---LINA-----PGSFTKCSNTSELKEFVNPSPLPREKEKLETVKYSNN-- 743
QY 732 SHHDDOYIAEAPTEHNGRKAOKKLTWEQKATYTN-----SPAATCGAOPFRGIAV 783
Db 744 -----AEDP-----KDL--MLSGERVLTQERSVSSISLVGFTYGTG-----ESI 783
QY 784 DLSTHWAGSSSNYSARQPVYAPLDRAVRAVNOYHANFPSTATMASCRCRRAGQ 843
Db 784 SLLEVSTLGR---AKTEP-----NKCVSOCALAEENPKGL--HGCGR--DNRNQTE 827
QY 844 VLVYKESMPATHLRMDPSTLASFPNYGTSSRNQMSQLHNSQYAHNOYKGSTSTSYG 903

Db 828 GFKY-----LGHEVNHRSRTSIEMESSELDAYLONTFKVRSQSPA 870
QY 904 SNLNGKIPITFEDLSRHQJLHRLRPHPRVYGLSLKKEIANWSENCGTOSQYKLG 963
Db 871 -----PFSNG-----NAEECATFSAHSGSL 892
QY 964 STGITSH 970
Db 893 KTKSKSH 899

RESULT 4

US-08-825-886-20

; Sequence 20, Application US/08825886
; Patent No. 5821328

GENERAL INFORMATION:

APPLICANT: KING, Mary-Claire

APPLICANT: FRIEDMAN, Lori

APPLICANT: OSTERMEYER, Beth

APPLICANT: ROWELL, Sarah

APPLICANT: LYNCH, Eric

APPLICANT: SZABO, Csilla

APPLICANT: LEE, Ming

TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN

TITLE OF INVENTION: CANCER

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HORBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/825,886

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/425,061

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-59563-3/DUB/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 494-8700

TELEFAX: (415) 494-8771

TELEX: 910 277299

INFORMATION FOR SEQ. ID NO. 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 900 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-825-886-20

Query Match 3.3%; Score 182; DB 2; Length 900;

Best local Similarity 18.4%; Pred. No. 1.2e-05;

Matches 200; Conservative 142; Mismatches 363; Indels 382; Gaps 49;

QY 48 VDPAQHCHSFIRGVAL--QKDKPKCSIRIFHQKKDEKHAASSPSS--YAKF 103
Db 31 IKPEVSTKCHIFCKMELKLNQKGPSCPLCKNDITKRSLOE---STRSQIWEEL 86
QY 104 RHWDCSKCLD----- 113

Db 87 LKICAFQDLDGTGLFYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRKRLLOSEPENPS 146
QY 114 -----KIKTSNDGTAPRTLPKAKONGT-----SQCSTTFVRSFTVPASVSOK 156
Db 147 LQETSLSVOLNIGTV-RLTKRKQRIQPKTSYVIELGSDSEEDTKATY--CSVDQE 203
QY 157 VSPSTOSSQGNKADRSTLPKSVQEGNDSKCNAPSGKNAAE-ANTDSPMKDLQPAQNYD 215
Db 204 LLQIT-----PGTRDEISLDAKKAACCFSTEDVTNTEHHQPSNN-- 244
QY 216 VAANVEDNTSDVYALPEVQITWHIEVNGADQPPSTPKLSEVYLKRNEDENGTETEL 275
Db 245 -----DLNTT-----EKRAEHRP-----EKYQSSVSNL 269
QY 276 VABQCN-----LTKDPMPMSGKERDQVABQCNLTQPKPVSQOK----- 314
Db 270 HVEPGTNTHAASSLOHENSLLTKD-----RANVEAEACNMSKQDGLARSHQNRAG 323
QY 315 CEQICNEPCEEVVLKRSKSKRTD---KILMKRQHSKRRTAQADVDAKLCRRKPK- 370
Db 324 SKETCND-----RRPTSEKTYDLNADPLCEKREMNKKLP-----CSENPDT 367
QY 371 -----VILLSEITIANQVEDSRDEVAHENAADCEDDRTIIPVPMESMDI-----PVS 420
Db 368 EDVPMITLNSSTQKVN-WEFSRDEL---LQSDSHDESSNAKVADVLDVILNEVDEYS 423
QY 421 NHTVGEDGLKSKN-----KTRKRYSDVDGSSILMNLNKKRRTGSVHHTVAH----- 470
Db 424 GSSEKIDILADPHALLCKSERVSKSV-SENEDLIFKTYRKAKASLNLSHVETNL 481
QY 471 -----PAGNLKSKVTPASTQHDHDENGDENTNGDITNKTQVC-QHSEI 514
Db 482 IIGAFVTEPQIOERPLTNKLRKRRTPSGLHPED-----FIKADLAQOKPEM 531
QY 515 ---STQRCSSKGTAGLKGKTHSASTRKYGESSTRNGONIHVLSAEDOCMETENSYS 571
Db 532 INQGNQTEQNGQVNT-----TNSGHEKTKGDSIQNKKNRPLES-----LEKESAKT 582
QY 572 HSAKVPAAEHDIQMSDLHEQSLPKRRKKQKLEVREKQTMIDIPMDIVELLAKNHER 631
Db 583 KAEPISSTISNNELINHSKAPKKNRLRRKSTRH-----IHAELEIVSRNLSP 633
QY 632 QLMTE---TDGSDIRIQSK-----TTADDCVVAAKDSDVYASVFDNISOOKSLAS 682
Db 634 PNCTELQIDSCSSSEIKKKKYNQMPVRHSRLQIMEGKEPATGAKKSNREQTSKHD 693
QY 683 OSTOKELQHLALTTOESHPQNF---OSTOQOQTHL-----RMEEMVTIAASSPLF 731
Db 694 SPTFPELK---LTNA---PGSFTKCSNTSELKEFVNPISLPREKEKELETVKYSNN-- 743
QY 732 SHHDDQYLAEPTEHMGKDKAKKLTWEQFKATRN-----SPAATCGAOFRCIQAV 783
Db 744 -----AEDP-----KDL-MLSGERYLOTEREVSESSISLVGTGYO---ESI 783
QY 784 DLTSTHWGSSSNVARSOPVIALDRVAREAVNOYHARNFPTIATWASKLCDRRNAG 843
Db 784 SLEEVSTLCK---AKTEP-----NKCVSOCAAPENPGL-IHGSK--DNRNTE 827
QY 844 VVLYPKESMPATHLLRMMDPSTLASFPNYGTSRRNOMESQLHNSQYAHNQYKGSYSTYG 903
Db 828 GFKY-----LGHEVNHRSRTSIEMESSELDAYLONTFKVRSQSPA 870
QY 904 SNLNGKIPITFEDLSRHQJLHRLRPHPRVYGLSLKKEIANWSENCGTOSQYKLG 963
Db 871 -----PFSNG-----NAEECATFSAHSGSL 892
QY 964 STGITSH 970
Db 893 KTKSKSH 899

RESULT 5

US-08-603-753D-2

Sequence 2, Application US/086037530
Patent No. 5891857
GENERAL INFORMATION:
APPLICANT: HOLT, JEFFREY T.
APPLICANT: JENSEN, ROY A.
APPLICANT: PAGE, DAVID L.
APPLICANT: KING, MARY-CLAIRE
APPLICANT: SZABO, CSILLA I.
APPLICANT: JETTISON, THOMAS L.
APPLICANT: ROBINSON-BENTON, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
STREET: BOULEVARD
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch, 800 KB storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: ductal carcinoma in situ, invasive
CELL TYPE: breast cancer and normal breast tissue
CELL LINE: not derived from a cell line
ORGANELLE: no
IMMEDIATE SOURCE:
LIBRARY: cDNA library derived from human
CLONE: obtained using published sequence
POSITION IN GENE:
CHROMOSOME/SEGMENT: unknown
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: BRCA1 protein
LOCATION: 1 to 1863
IDENTIFICATION METHOD: observation of mRNA and
IDENTIFICATION METHOD: antisense inhibition of BRCA1 gene

OTHER INFORMATION: BRCA1 protein has a negative
OTHER INFORMATION: regulatory effect on growth of human mammary cells.
PUBLICATION INFORMATION:
AUTHORS: Miki, Y., et. al.
TITLE: A strong candidate gene for the breast and
TITLE: ovarian cancer susceptibility gene BRCA1.
JOURNAL: Science
VOLUME: 266
PAGES: 66-71
DATE: 1994
RELEVANT RESIDUES IN SEQ ID NO: 2: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 1214-1223
US-08-603-753D-2
Query Match 3.3%, Score 180, DB 2: Length 1863;
Best Local Similarity 17.9%, Pred. No. 5.2e-05;
Matches 221; Conservative 174; Mismatches 387; Indels 452; Gaps 62;
48 VDEPAHOCEHPSIRGVYALL--OKKDPKFCSLRIFHDQKCDDEHKASSPFS--VAKF 103
Db 31 IKEPVSTKCDHIFCKFCMLKLNQKQSPCLCANDITKRSLOE-----STFSLQVEEL 86
Qy 104 RRMDCSKCLD----- 113
Db 87 LKIIAFOPLDGLGVANSYFAKKENNSPEHLKDEVSIIQSMGYNNRAKRLQSEPNPS 146
Qy 114 -----KKTSPNGTAPRTLPKQNGT-----SDGCTFFVSTFPAPVSGSK 156
Db 147 LOETSLSVOLSNIGTV--PRLRTKRIQIQOKTSVYIELSDSSSEDPVNNATY--CSVGDOE 203
Qy 157 VSPSTQSSGKNADRSTLPKSVQEGNDSKCNAPSKNGAAE--ANTDSPKDLQGPANVD 215
Db 204 LIQIT-----PGSTRDEISLDSKAKACESEFDVYTHRHQPSNN-- 244
Qy 216 VAANVSEDNTSVYDGLBEVPO-----ITWHIEVNGADPPSTPKLSEVYLKRNEDN 268
Db 245 -----DLNTEKRAAEHNPKEYQSSVSNLHVEPCGNTNASSLQHNSSLTLTKDRM 297
Qy 269 GKTEETLVAEOCNLTKDRPNPSGKERDQ-----VAEOCNLTKDRKPVYSGQR-----CEQ 317
Db 298 NVEK----AEFCNKSQKOP-----GLARQHNRRWAGSKETCN--DRTSTPTEKRVLDNDP 346
Qy 318 ICNE-----PCEE-----VYLKRSKSKR-----KTDKKMKKQOHSKRRQAQ 355
Db 347 LCEKKEWKKQKLPCESENPROTEDEVPITLNSIQVNEKFSDELGSDDSHDESESN 406
Qy 356 ADVSDAKLCRRK-----PKKVRLLSEIINANQVEDSKRDEVRHNADPCED----- 402
Db 407 AKVADVLVLANVDEYSGSSEKIDLLAS--DPHEALICKSDVHAKSVESDIEDKIFGKT 464
Qy 403 --DRSTIVPMKESMDI-----PVSNH-----TVG---EDGLKSKNK 435
Db 465 YRKKASLPNLSHVTENLLIGAFVSEPOIIQERPTNLKLRKRKRPSTGLHPEDFIKKA-DL 523
Qy 436 TKRKYSADVVDGSS-----LMNNLN--GKKKRT-----GSHVHTV 468
Db 524 AVQKTPPEMINTOSTNOTEDONGOVAMNITSGHEKTKTGDSIQNEKNPNPIESLEKESAFRTK 583
Qy 469 AHP--AGNLSNKKVPTASTOHDDEDTENGL-----DTFMHKTQVC-----QHVSEIS 515
Db 584 AEPISSTISNEL--ELNIMHNSKAPKKNNRLKRSSTRIRIHALVELVSNLSPPCTELQ 640
Qy 516 TORCSSKGTAGLSKGTHTSAATYGGESTRNGONIHVLSMED-----QCOMETNSV 569
Db 641 IDSCSS-----SEIIRKKRYNOMPVHSHNLQIMEKKEPATGAKKSNKKNPEQFS 689
Qy 570 LSHSAKVPAPENDIOI-----MSDLAE---OSLPKKKKKKKQLEYVRREKQTMID 614
Db 690 KRHSDTFP---ELKLTNAPGSPFTCSNTSELKEFVNPSPLEKEKEKLEYKAVSNNAED 746
Qy 615 DIPMDIVELLAKNQHROLMETDCSDIRIOSKTTADDCIVYAKGSDY----- 666

```

Db 747 --PKDLM-----LSGRVLOTE-----RSVSSS-----ISLVPGTDYGTQESISL 785
QY 667 --ASSVFDNISOOKSLASSTOKELQHL-----ALTQESPHPOFGSTQEQOHL 716
Db 786 LEVSTIGRAKTEPNKCVSOGAFENPKGLIHGCKDNRRDTEGFKPLGHEVNHSHRETSI 845
QY 717 RMEWMTIASASLFEHHDOYIAE-----AP-----TEHWG--RK 750
Db 846 EME-----SELDAQYLONTFKYKSKROSFAPFNPNGNAEBCATFFSHSGSLK 894
QY 751 DAKKLME-----QFKATRNSPAAT-----CGAOPRPGIOAVDLSTHWGSS 794
Db 895 QSPKVFEECEQKRENGKKNESNIKPVQVNTAGFPVYQKDKP-----VDNACSIKGS 950
QY 795 SNAASQ-----PYAPLDRIAE--RAVNOVARNPSTIATM 830
Db 951 RCLSSQFRGNETGLTPPKHGLLOPNRYRIPLEPIKSTVKTKCKKNLEEN----- 1003
QY 831 EASKLCDRRNAGOVLYPKESMPATHLNMMDSTLASIPNYGTSRNOMESQLHNSQYA 890
Db 1004 EESMSPEREMG-----NENIPST-----VSTIRNNIRENVEKEASS 1041
QY 891 HN-QYKGSTSTSYGSLNKGKIPLTFE---DISRHQLHDLHPLRPHPRVGLSILQRE 945
Db 1042 SNINEYGSSSTNEVGSSIN-EIGSSDENIQAEIGRNGPILNMLR-----LG-VLOPE 1092
QY 946 IANWSENGCQSGYKLGVSIGTISHOMNREHRE 979
Db 1093 V-----YKOSLPESNCKHPEIKQEYE 1114

```

RESULT 6

US-09-099-753-2
Sequence 2, Application US/09099753

Patent No. 6149903

GENERAL INFORMATION:

APPLICANT: HOLT, JEFFREY T.

APPLICANT: JENSEN, ROY A.

APPLICANT: PAGE, DAVID L.

APPLICANT: KING, MARY-CLAIRE

APPLICANT: SZABO, CSTILA I.

APPLICANT: JETTON, THOMAS L.

APPLICANT: ROBINSON-BENTON, CHERYL L.

APPLICANT: THOMPSON, MARILYN E.

TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2

TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON

TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARLES A. TAYLOR, JR.

STREET: SUITE 1401, UNIVERSITY TOWER, 3110 TOWER

STREET: BOULEVARD

CITY: DURHAM

STATE: NORTH CAROLINA

COUNTRY: USA

ZIP: 27707

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 800 KB storage

OPERATING SYSTEM: Windows 3.1

SOFTWARE: WORD PERFECT 6.1 and ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/099,753

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/603,753

FILING DATE: 20 FEB 1996

APPLICATION NUMBER: U.S. 08/373,799

FILING DATE: 17 JAN 1995

ATTORNEY/AGENT INFORMATION:

NAME: ARLES A. TAYLOR, JR.

REGISTRATION NUMBER: 39,395

REFERENCE/DOCKET NUMBER: 1242/2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 493-8000

TELEFAX: (919) 419-0383

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1863

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: protein

HYPOTHETICAL: no

ANTI-SENSE: no

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE: adult

TISSUE TYPE: female breast

CELL TYPE: ductal carcinoma in situ, invasive

CELL TYPE: breast cancer and normal breast tissue

CELL LINE: not derived from a cell line

ORGANELLER: no

IMMEDIATE SOURCE:

LIBRARY: cDNA library derived from human

CLONE: obtained using published sequence

POSITION IN GENOME:

CHROMOSOME/SEGMENT: unknown

MAP POSITION: unknown

UNITS: unknown

FEATURE:

NAME/KEY: BRCA1 protein

LOCATION: 1 to 1863

IDENTIFICATION METHOD: observation of mRNA and

IDENTIFICATION METHOD: antisense inhibition of BRCA1 gene

OTHER INFORMATION: BRCA1 protein has a negative

OTHER INFORMATION: regulatory effect on growth of human mammary cells.

PUBLICATION INFORMATION:

AUTHORS: Miki, Y., et. al.

TITLE: A strong candidate gene for the breast and

TITLE: ovarian cancer susceptibility gene BRCA1.

JOURNAL: Science

VOLUME: 266

PAGES: 66-71

DATE: 1994

RELEVANT RESIDUES IN SEQ ID NO: 2: granin box

RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 1214-1223

US-09-099-753-2

Query Match

Best local similarity 17.9%; Pred. No. 5.2e-05;

Matches 221; Conservative 174; Mismatches 387; Indels 452; Gaps 62;

```

QY 48 VDEPAOHCHEFSIRGVALL--QKDKPFCSSRIFHOQKDEKASSPSS--YAKF 103
Db 31 IKEPVSKCDHIFKCMKMLNOKKGPSCPLCKNDITKRSIQE---STRSOLVEL 86
QY 104 RHMDSKCD-----SDGCSITFVRSTFVPAVSGSK 156
Db 87 LKTIICAFOLDTGLEYANSYFAKKENNSPEHLKDEVSIIOSKGYRRRAKRLQSEPNPS 146
QY 114 -----KLKTSNGTAPRLPAKONGT-----SDGCSITFVRSTFVPAVSGSK 156
Db 147 LQETSLSVQLSNLGYV-RLTRTKQRIQPKTSYIELGSDSSBDYVNAKATY--CSVGDQE 203
QY 157 VSPSTOSSGCKNADRSTLPKSVQEGNDKCNAGSKNGAAE-ANTOSPKNKDLQPPKOND 215
Db 204 LLOIT-----POGTRDEISDSAKKAACEFSETVTNTTEHHQPSNN-- 244
QY 216 VAANVSEDNTPVDVGLPEVPO-----ITWHIEVNGADQPPSTPKLSEVFLKRNEDEN 268
Db 245 -----DLNITTKRAERHPEKYGSSVSNLHYEPCGTHTASSLQHEHNSILLTKDRM 297

```


0Y	269	GKSETVLAEQCNLTQDPNPMSEKERO	-----VAEQCNLTTPDKPVSGOK	-----CEQ	317
Db	298	NVEK-----AEFCNNSKOP	-----GLANSOHNRNMGSKETCN	-----DRKTPSTEKVLDLNDP	346
0Y	318	ICNE-----PCE	-----VJLKRSSKSKR	-----KYDKLMMKQOOSKKRPATQ	355
Db	347	LCEKKNKQKLPCESENPRTEDVPWITLNS	IQKXNEMFNSKDELGSDDSDHDESESN	406	
0Y	356	ADVSADKLCRRK	-----PKVRLLEIIMANQVEDSDEVHRENA	PCD-----	402
Db	407	AKVADVLVDLVNDEVDYSGSEKIDLLAS	-DPHEALICKSDRVHKSXSVSDIEDTKFGKT	464	
0Y	403	-----DASTIPVMEVSMOI	-----PVSMH	-----TVG	435
Db	465	YRKASLCPNLSHVTENLLIGAEVSEPOIIQERPLTKLKRKRPRVSGLHPEDFIKA	-DL	523	
0Y	436	TKRRSPDVDDGSS	-----LNMWLN-GKKRRT	-----GSVHNHTV	468
Db	524	AVQCTPRMINGTQOTQONQVNNITNSGHENTK	QDSIQNEKNRPIESLEKESAFKTK	583	
0Y	469	AHP-AGLNSKKVTPTASTQHDDENDTENGL	-----DTNMKRTDYC	-----QHYSEIS	515
Db	584	AEPISSTISNEL	-----ELINIMHNSKAPRKRLRKRKSTRIHIALELVYSNNLSRPCTELQ	640	
0Y	516	TQRSSSGKGTAGLSKGTHTHASTKYGGBESTRNGQNIHVSLED	-----QCOMETENSV	569	
Db	641	IDOSS	-----SEBKKKKKNOMPVRRSRNLOLMEKEEPATGAKKSNKNEQOTS	689	
0Y	570	LSHSAKVPASPAHDIOI	-----MSDLHE	-----QSLPKKKKQKLEATREKQTMID	614
Db	690	KRHDSOTFR	-----ELKLTNAPGSEFTKCSNNTSELEEFVNPSPREKKEKLETYKVSNNAD	746	
0Y	615	DIPMDIYELLAKNOHEROLMETEDCSDIRNIOKQRTIADDCVIVAKQSDY	-----	666	
Db	747	-PRDLM	-----LSGEHVLQTE	-----RSVESS	785
0Y	667	-----ASSVEDTNSQOKSLASQSTOKELQOHL	-----ALTTQESPBPONFQSTQDOQOHL	716	
Db	786	LEVSTIGAKTEPRKCVSQAAPENPKGLIHGCSKDNRRDTEGFKYPLGHEVHNSHRETSI	845		
0Y	717	RMEEMVTIAASSPLFSHHDDQYIAE	-----AP	-----TEHNG	750
Db	846	EMEB	-----SILDQOYLQNTFKYKSKRQSPAFPSNPGNAEECATTEFAHSGSLAK	894	
0Y	751	DAKLITWE	-----QFKATTNRSPAT	-----CGAOFRRPGIOAVDLTSTHVWSS	794
Db	895	QSPVTEFECEQKEENQCKNESNIKPVQTVITAGFVWQOKQK	-----YDNMKCSIKGSS	950	
0Y	795	SNVYSRQ	-----PVIAPIIDRAE	-RAYNQVIARNFPTSIATM	830
Db	951	RECLSSQFRNGENTGLIPNKHGILLQNDPYRILPFLPIKSFVKTKCKKNKLEENF	-----	1003	
0Y	831	EASKLCQBRNAGQVLYLPKESMPATHLLRMDPSTLASFPVNTGSSRNOMESQILNSQYA	890		
Db	1004	EEHSMSEPERENG	-----NENIPST	-----VSTISRRNIRBNVKEASS	1041
0Y	891	HN-QYKGSTSTSYGSLNIGKPIPLTFE	-----DLSRHQJLHDLNRPRLRPNRVGVLGSLIOKE	945	
Db	1042	SNINEVGSSSTNEVGSSIN-ELIGSSDENIGAEIQRNNGRPLNMLR	-----LG-VLOPE	1092	
0Y	946	IANNSEKQGTOSQKGLCVSTGITSHOMKKEHFE	979		
Db	1093	V-----YKQSLPSNCKNHPKIKQOJE	1114		

RESULT 7
US-08-986-106-2
; Sequence 2, Application US/08986106
; Patent No. 6177410
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.

```

1 APPLICANT: KING, MARY-CLAIRE
2 APPLICANT: STEINER, MITCHELL S.
3 APPLICANT: ROBINSON-BENTON, CHERYL L.
4 APPLICANT: THOMPSON, MARILYN E.
5 TITLE OF INVENTION: THERAPEUTIC METHODS FOR
6 TITLE OF INVENTION: PROSTATE CANCER
7 NUMBER OF SEQUENCES: 26
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: ARLES A. TAYLOR, JR.
10 STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
11 STREET: BOULEVARD
12 CITY: DURHAM
13 STATE: NORTH CAROLINA
14 COUNTRY: USA
15 ZIP: 27707
16
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
19 COMPUTER: IBM PC/XT/AT compatible
20 OPERATING SYSTEM: Windows 3.1
21 SOFTWARE: WORD PERFECT 6.1 and ASCII
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/986,106
24 FILING DATE:
25 CLASSIFICATION:
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: 08/603,753
28 FILING DATE: 20 FEB 1996
29 ATTORNEY/AGENT INFORMATION:
30 NAME: ARLES A. TAYLOR, JR.
31 REGISTRATION NUMBER: 39,395
32 REFERENCE/DOCKET NUMBER: 1242/3
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: (919) 493-8000
35 TELEFAX: (919) 419-0383
36 INFORMATION FOR SEQ ID NO: 2:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 1863
39 TYPE: amino acid
40 STRANDEDNESS: single
41 TOPOLOGY: unknown
42 PUBLICATION INFORMATION:
43 AUTHORS: Mikl, Y., et. al.
44 TITLE: A strong candidate gene for the breast and
45 TITLE: ovarian cancer susceptibility gene
46 TEST TITLE: BRCA1.
47 JOURNAL: Science
48 VOLUME: 266
49 PAGES: 66-71
50 DATE: 1994
51 RELEVANT RESIDUES IN SEQ ID NO: 2: granln box domain
52 RELEVANT RESIDUES IN SEQ ID NO: at amino acids 1214-1223
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

```

OY 216 VAANVEDNTSVGVGALPEVPO-----ITWIEVNGADOPSPKLSSEVLKRNEDEN 268
DB 245 -----DLNTEKRAAERHPEKYOSSVSNLHVPCGTITHASSLOHENSLLITRDM 297
OY 269 KTEETLVAEOCNLTKDPNPMGKERDO-----VAEOCNLTKDPKPVSGOK-----CEO 317
DB 298 NYEK-----AEPCNSKQOP-----GLARSOHNWAGSKETCHI---DRPRPSTFKKYDLNADP 346
OY 318 ICNE-----PCPE-----VILKSSSKSR-----KTDKLLMKKQOHSKKRTAQ 355
DB 347 LCEKEMWKKOKLPCSENPRDTEVPWITLNSICKVNMWPSRSDLLGSDSHGSESEN 406
OY 356 ADVSDAKLCRRK-----PKVRLSEITIANOVEDRSDEVHRENADPCED-----402
DB 407 AKVADVILVLENEVEYSSSEKIDLLAS--DPHEALICKSDRVHKSVESDIEDKIFGKT 464
OY 403 --DRSTIVPEVMSDI-----PVSNN-----TVG--EDGLKSSKNK 435
DB 465 YRKKASLPLNSHVETNLLIGAFVSEPOILIOERPLTNKLRKRRPTSGLHPEDFIKKA-DL 523
OY 436 TKRKSDVVDGSS-----LMNMLN-GKKKRT-----GSVHTV 468
DB 524 AVOKTPEMINGTNOTONGOVAMNITNSGHEKTKGDSIONEKNPNPIESLEKESAFKTK 583
OY 469 AHP-AGNLSNKKVPTASTOHDDENDTENGTL-----DTNIAKTQVC-----OHVSEIS 515
DB 584 AEPISSTISNEL-----ELNIMNSKAPKKNRLRKRSSTRHITALELYVSRNLSPPCTELO 640
OY 516 TORSSSKKTAGLSKGTNHAASATYGESESTRNGONIHVJSAED-----OCOMETENSV 569
DB 641 IDSCS-----SEELIKKKKNOMPVHSHRNLOJMGKEKPEATGAKKSNKPFNEQTS 689
OY 570 LSHSKVSPAEHDIOI-----MSDLHE--QSJPKKKOKLEVTREKQTMID 614
DB 690 KRHDSDTP--ELKLTNAPGSFTWCSTSEKFEVNDSPREEKEKLELYKVSNNAD 746
OY 615 DIPMDIVELLAKNOHEROLMETEDCSINRLOSKTADDJCIVIAKAGSDY-----666
DB 747 --PKOLM-----LSGERVLOTE-----RSVSSS-----ISLVGTIDYGTQOESTSL 785
OY 667 --ASSVFDTNQOKSLASOSTOKELQHL-----ALTOESPHPQNFQSTOEOQTHL 716
DB 786 LEVSTLGAKTPEPKCVSQCAFENPKGLIHGCSKDNENRDEGFKYPLGHEVNHGRETSI 845
OY 717 RHEWVTTAASSPLFSHDDOYIAE-----AP-----TEHNG--RK 750
DB 846 EMEE-----SELDAOYLONTFVSKRSQSPAPPSJPGNAEECATPSAHSGLSK 894
OY 751 DAKKLTWE-----QKATTRNSPAT-----CGJQFRPGIOAVDLITSTHVMGSS 794
DB 895 QSPKTYTEEOKEENOGKNESNIKPVQYVNTIAGFPVVGJXKXP--VDNAKCSIKGGS 950
OY 795 SVYASRO-----PYIAPLDKRYAE-RAVNOYHARNFESTIATM 830
DB 951 RPLCSOFRKNGTGLITPNKHLQONPYRIPLPIKSPYKTKCKKNLLENF-----1003
OY 831 EASKLCDRNAGQVLYLPRESMPATHLLRMDPSTLASPYNATGSSRNOMESQLHNSOYA 890
DB 1004 EHSHPSPREKNG-----NENIPST-----VSTISRNKIRNVKKESS 1041
OY 891 HN-OYKSTSTSYGSLNGKIPLFE-----DLSRHQLDJBHRLPHRPVGLSILQKE 945
DB 1042 SHINEGSSSTNVESSIN-EIGSSDENQIAELGRNRPKNANMLR-----LG-VLQPE 1092
OY 946 IANMSENCGTQSGYKLVSTGITSQOMNKEHFE 979
DB 1093 V-----YKOSLFGSNCKHPEIKQOYE 1114

```

RESULT 8
 US-09-007-678B-49
 ; Sequence 49, Application US/09007678B

```

; Patent No. 6342483
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: OBERMILLER, PATRICE S.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: METHOD FOR DETECTION AND TREATMENT OF BREAST CANCER
; FILE REFERENCE: Attorney Docket No. 6342483 1242-1-2-2
; CURRENT APPLICATION NUMBER: US/09/007,678B
; PRIOR FILING DATE: 1998-01-15
; PRIOR APPLICATION NUMBER: 08/373,799
; PRIOR FILING DATE: 1995-01-17
; PRIOR APPLICATION NUMBER: 08/182,961
; PRIOR FILING DATE: 1994-01-14
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Microsoft Wordpad
; SEQ ID NO 49
; LENGTH: 1863
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc:feature
; LOCATION: (1472)
; OTHER INFORMATION: Xaa-any amino acid
US-09-007-678B-49

Query Match 3.38; Score 180; DB 4; Length 1863;
Best Local Similarity 17.98; Pred. No. 5.2e-05;
Matches 221; Conservative 174; Mismatches 387; Indels 452; Gaps 62;

OY 48 VDEPAHOCHEFSIRGVALL--OKDPKFCSLRIFHDOCKDEHKASSPFS--YAKE 103
DB 31 IKEPVSTKCDHIFKFCMLKLNKKGPSCPLCKNDITRSLOE-----STRSQLYEEL 86
OY 104 RRMDCSKLD-----113
DB 87 LKTIARFQLDIGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRKRLLOSEPENS 146
OY 114 -----KIKTSNCTAPRTLPKQNGT-----SDGCTTFPRSTFVPVPSVSOK 156
DB 147 LOETSLVSQSLNIGTV--RLTEKTKORIQPKTSVYLEGSDSSSEDTVKATY--CSVGDQ 203
OY 157 VSPFQSSQGNMARNRSTLPKSVQEGNSKCNAPSGKNGAAE-ANTDSPMKDLQPAONYD 215
DB 204 LLQIT-----POGTDEISLDSAKKAACEFSTDTVTNTEHHQPSNN--244
OY 216 VAANVEDNTSVGVGALPEVPO-----ITWIEVNGADOPSPKLSSEVLKRNEDEN 268
DB 245 -----DLNTEKRAAERHPEKYOSSVSNLHVPCGTITHASSLOHENSLLITRDM 297
OY 269 KTEETLVAEOCNLTKDPNPMGKERDO-----VAEOCNLTKDPKPVSGOK-----CEO 317
DB 298 NYEK-----AEPCNSKQOP-----GLARSOHNWAGSKETCHI---DRPRPSTFKKYDLNADP 346
OY 318 ICNE-----PCPE-----VILKSSSKSR-----KTDKLLMKKQOHSKKRTAQ 355
DB 347 LCEKEMWKKOKLPCSENPRDTEVPWITLNSICKVNMWPSRSDLLGSDSHGSESEN 406
OY 356 ADVSDAKLCRRK-----PKVRLSEITIANOVEDRSDEVHRENADPCED-----402
DB 407 AKVADVILVLENEVEYSSSEKIDLLAS--DPHEALICKSDRVHKSVESDIEDKIFGKT 464
OY 403 --DRSTIVPEVMSDI-----PVSNN-----TVG--EDGLKSSKNK 435
DB 465 YRKKASLPLNSHVETNLLIGAFVSEPOILIOERPLTNKLRKRRPTSGLHPEDFIKKA-DL 523
OY 436 TKRKSDVVDGSS-----LMNMLN-GKKKRT-----GSVHTV 468
DB 524 AVOKTPEMINGTNOTONGOVAMNITNSGHEKTKGDSIONEKNPNPIESLEKESAFKTK 583

```


QY	561	QÇQTEHNSVLSHAKSVAPAHDIOT-----MÇJLHE-----QSLPKKKKROKLEY	605
Db	681	SNKNEQTSKRHSDTFR---ELKLTNAPCSFTKCSNITSJLKEFVNPSPJPREKEKEKLET	737
QY	606	TREKOTMIDDIPIMDIYELLAKNOHEROLMETETDCSDINRIGSKTTADDDCVIAAKGSD	665
Db	738	KVAVSNNAED---PKJLM-----LSGERLYOTE-----RS/JESSSS-----ISLVPGTD	776
QY	666	Y-----ASSVPEPTNSOOKSLASOSQKELQGHJL-----ALTOJESPHQNFQ	707
Db	777	YGTQESTLSLEVTSLTGAKTEPKVCYSQCAFEFNPGLJHCSKDNDRNDBEFKTPJLGH	836
QY	708	STOBQOQTHLRMEENVYTTAASSPLFSSHDDQYIAE-----AP-----	743
Db	837	VNHRETSIEEMEE-----SELDAQYLQNTFKVASIQSFAFESNPGAEEECATF	885
QY	744	TEHMG---RKDAKKLTWE-----QKATTRNSPAT-----CGAQFRRPGQAVDL	785
Db	886	SAHSGSLKKSOPKAPTECEQKEENOGKESNIRKPVQVNTITAGFPVVGQKDP---VDN	941
QY	786	TSTHVMSSSNVYASRQ-----VYAPLDRYAE---RAVNOVBAR	821
Db	942	AKCSIKGGSRRCLSSQCRNGMETGLITPNKHGILLQNYRI2PLEFPIKSYFKCKCKNNILEE	1001
QY	822	NFPSTIATMEASKICDRNRNAGOVYLYPKESMPATHILIRM4DDSTLASFPNTGTSRNOHE	881
Db	1002	NE-----EHSNSPBRERG-----NENIPST-----VSTISRNMR	1032
QY	882	SQJLHNSQYAHN-QYKQSTSTSYGSLNGLKJPLTFE-----JLSRHOJDLHRLPRLPVRG	936
Db	1033	ENVEFKKASSNININVGSTJENWGVSSIN-ELGSSDENIQA2LGNRNPRIKJLAALR-----	1081
QY	937	VJGSLLOKELAINSENCGTOSGYKLGAVSTGITSIHOMNRK3HEF	979
Db	1086	-LG-VLOPEV-----YKQSLPGSNCNKHPIKQJEVE	1114

RESULT 10
US-08-825-886-22
Sequence 22, Application US/08825866
Patent No. 5821328
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HORBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825.886
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/425,061
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, Richard A

1 REGISTRATION NUMBER: 36, 627
 2 REFERENCE/DOCKET NUMBER: A-5963-3/D/B/RAO
 3 TELECOMMUNICATION INFORMATION:
 4 TELEPHONE: (415) 494-8700
 5 TELEFAX: (415) 494-8771
 6 TELEX: 910 277299
 7 INFORMATION FOR SEQ ID NO: 22:
 8 SEQUENCE CHARACTERISTICS:
 9 LENGTH: 1202 amino acids
 10 TYPE: amino acid
 11 STRANDEDNESS: single
 12 TOPOLOGY: linear
 13 MOLECULE TYPE: protein
 14 US-08-825-886-22

Query Match	3.2%	Score 179;	DB 2;	Length 1202;
Best Local Similarity	17.9%	Pred. No. 3.1e-05;		
Matches 222;	Conservative 170;	Mismatches 381;	Indels 470;	Gaps 62;

QY	48	VDEPAQHOCHEFSIRGVALL--OKXPKPCSLRIFHODKKODEKASSPPS--YAK	103
Db	31	IKEPSTXCDHIEFKFCMLKLLNOKKSPSOCPLCKNDITKRSIQE--STRESQVEEL	86
QY	104	RRMDCSKLD-----	113
Db	87	LKIIICAPLODGLGVANSYNPAKKENNSPEHLDEVSIIISMGIYRNKAKLLQSPENPS	146
QY	114	-----KIKTSIDNGTABPTLEPAKONGT-----SDGCSITFEYKSTFVPASVGSOK	156
Db	147	LOETSLASVOLSNLGTV--RLTLTKQRIORPOKTSYVIELGSSSEBTPYAKATV--CSYGDOE	203
QY	157	VSPSTQSSQGNNAARSTLPKSVQSGNDSKCNAPSGKKGADE--ANTQSPMDLOGPQMYD	215
Db	204	LIQIT-----PGGTDEEISLSBAAKACESEBTDTYNTENHOPSN--	244
QY	216	VAANVSEDNTPSDVGCALPEVPO-----ITWHLIENGADQPESTRKLSYEVULKREDEEN	268
Db	245	-----DLNTEKRAABERHEKXQGSVSLSLHNEPCTGTHMSSIQHENSLSLLTKDRM	297
QY	269	GKTEBTLVAEOCNLTKDPNPMSGKEBPO-----VAEOCNLTDPKPYSGOK-----CEO	317
Db	298	NVEK-----AEFCNKSOKP--GLARSOHNRMAWASKETCN--DRRTPESTKKVDINADP	346
QY	318	ICNE-----PCBE-----VILKRSKSKR--KTDPKILMKQOQSKKRTAQ	355
Db	347	LCERKENMKOKPLPCESENPROTEDYPIWTLNNSIOKVENWERSDELLGSDSHDBEESN	406
QY	356	ADVSD-----AKLCRRKPKRYLLSEIIANQVEDSRS	388
Db	407	AKYADVLDVLEWVEDEYSSSEKIDLDSPHEALIC-----KSERVHSHKVSYESN--IEDKIF	461
QY	369	DEVHRENAADCEDDRSTIIPYPM-----EYSDIPIVSNH-----TVG--EDGL	429
Db	462	GKTYRKKKASLP--NLSHVLENLIGAFVTEPOLIOERPLTNLKKRRKRPSTGLHPEDFI	518
QY	430	KSSKKKTKRKXSDVDQSS-----LMWLN--GKKKRTGSYVHTGAHPAGNLSNK	478
Db	519	KKA-DLAVQKLPENINOSTNOTEDONGVMMITNSGHNENKTKGDSIQNEKKNPNIIESLEKE	577
QY	479	-----KYTPASTOCHDDENDTENGDLTNMKTUV-----COHV-----	511
Db	578	SAFKTKAPRISISSIMNE-----LELNHNSKAPKKNRLRRKSSSTRHIALBELVYSRNL	631
QY	512	-----SELSTORCSKSGKTAGLSKGKTHSASTYGESESTRNGONIHVLSAED-----Q	560
Db	632	SPPNCTELQIDSCSS-----SEEEKKKYUNOMPRVHSNINLOLMGKRPATGAKK	660
QY	561	COMETENSVLHSHKVPSPAEHDIOI-----MSDINE-----OSLKKKKKKOLEV	605
Db	661	SNKPRBJTSKRHDSDFP--ELULTNAPSGFTYCSMTSELKEPVNSLPREREKEKLET	737
QY	606	TREKOTMTIDDPMDIVELAKNOHEROLMETDSDINRIQSKTTADDCVIAAKDSD	665


```

Db 837 VNHRSRTSIEMEE-----SELDAYIANTFKVSKROSFAPSPNGNAEECATF 885
OY 744 TEHWG--RKDAKLTWE-----QFKATRNSPAAT-----CGAOFREGIOAVDL 785
Db 886 SAHSGSLKROSPRVECECHKEENCKNESNIKPVOTVITAGFPVVGKDKR-----VDN 941
OY 786 TSTVWSSSNVARSQ-----PVAPLDRYAE--RAVNOVHAR 821
Db 942 AKCSIKGSRFLCSOFKRNENGLITPKHGLIÖNDYRIPPEPIKSFVTKCKKNLLEE 1001
OY 822 NPPSTIATNEASKLCDRRNAGOVLYPKESMPATHLLRMDESTIASFPNYGTSSNOME 881
Db 1002 NF-----EEHSMSEEREMG-----NENIPST-----VSTISNNNR 1032
OY 882 SOLHNSOVANH-QYKGSTSTSGSNLNGKIPLTFE---DLRHOHLHLRPLRPHRPG 936
Db 1033 ENVFKEASSNNINEYSSSTNEVSSSTN-ELGSSDENIOLELGRNRGPKLNAMLR----- 1085
OY 937 VGLSLÖKEIANMSENCGTSGYKLGIVSTGITSHOMNRIEHE 979
Db 1086 -LG-VLOPEV-----YKOSLPGSNCKHPEIKIÖYE 1114

```

RESULT 12

```

US-08-825-886-23
; Sequence 23, Application US/08825886
; Patent No. 5821328

```

GENERAL INFORMATION:

```

APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Mung
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,886
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/425,061
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

```

US-08-825-886-23

Query Match 3.2%; Score 179; DB 2; Length 1363;
 Best Local Similarity 17.9%; Pred. No. 3.8e-05;
 Matches 222; Conservative 170; Mismatches 381; Indels 470; Gaps 62;

```

OY 48 VDEPAHOCEHFSIRGVALL--QKKDPKCSLSRIFHDQKDEHKASSPFS--VAKF 103
Db 31 IKEPVSTKCDHIFKFCMKLKNÖKKGFSQCPCKNDITKSLÖE---STRFSQVLEL 86
OY 104 RMDQSKCD----- 113
Db 87 LKIIAFOÖDPTÖLEYANSYNFAKKENNSPEHLKDEVSIIÖSGYNNRAKRLÖSEPNES 146
OY 114 -----KLKTSNGTAPRTLPAKÖNGT-----SDGCSITFVRSFPVAVSGÖK 156
Db 147 LÖETSLSVQSLNIGIV-RLTRFKÖRIÖPKTSVYIELSSSESDTVNKATY--CSVGÖE 203
OY 157 VSPSTÖSGKNAADSTLPKSVÖEGNDKCNAPSGKNGAAE-AMTDSPKKDÖGPAÖNYD 215
Db 204 ILÖIT-----PÖGTREIÖSLDSAKKACEFSEÖTVNTEHHÖPSNN-- 244
OY 216 VAANVEDNTSVYÖGALPEVPO-----ITWHEVNGADÖPSPPKLSEVYLKRNEDEN 268
Db 245 -----DLNTEKRAAHERHPEKYÖGSSVSNLHVPCÖGTNHASSLÖHENSLLTKÖRM 297
OY 269 GTEETLVAÖCÖNLTKDPNPSGKERDÖ-----VAÖCÖNLTKDPKPYSGÖK-----CEQ 317
Db 298 NYEK-----AEFCNKSÖKÖP-----GLARÖHNRWAGSKÖTCN--DARTSTÖKÖVLDNADP 346
OY 318 ICNE-----PCÖE-----VYLKRSÖKÖR--KTDKLLKÖQÖHÖSKKRAQ 355
Db 347 LÖERKEMKÖKLPÖSENPRDÖEDVWITLNSIÖQVNWEMFSRSDLLGSDSDHÖSESÖN 406
OY 356 ADVSD-----AKLÖRÖPKVRLSEIIANÖVEDERS 388
Db 407 AKVADVLDVNEVDEYSSÖSEKIDLLASDPHÖLIC-----KSERHVSÖSVÖSN-IEDÖIF 461
OY 389 DEVENADADCEÖDRSTIPPM-----EVSMDIPVSNH-----TVG--EDGL 429
Db 462 GÖTYKÖKSLP--NLÖSVTENLIIGAFVPEÖIÖEPÖLNKLRKÖRÖPSGHLPEÖFI 518
OY 430 KSKÖNÖTKRKYSDVYDÖSS-----LMNÖLN--GKKRÖTSVHHTVHPAGNÖSNK 478
Db 519 KKA-DLAVÖKÖTPÖMÖNGÖTÖEQNGVYMNÖTNSGHEKÖTGDÖIÖNÖKNPÖPLESÖKE 577
OY 479 -----KÖTPÖASTÖHDÖENDÖENGLDÖTMKÖTDV-----CÖHV----- 511
Db 578 SAFÖTKAÖPISSISÖNME-----LELNHNSKAPKÖRNLRRKSÖSTRIHÖLELVVSÖHL 631
OY 512 -----SEIÖTÖRCSÖSKÖTAGLSKÖTÖHSAÖSTYÖGÖSTRNGÖNIHLSÖAD-----Q 560
Db 632 SPNÖCTELÖIDÖSCSS-----SEIÖKKÖKIYÖNÖKPYHNSÖNLÖMEKÖEPÖATGÖAKK 680
OY 561 CÖMETENSÖVLSHÖKAVSÖPÖHDÖIÖ-----MSDÖLH--ÖSLPÖKKÖKÖLEV 605
Db 681 SNKPNÖÖTÖSKRHDÖDÖTP--ELKÖITNAPGÖFTÖCSNÖSEIÖKEFVNSÖLÖPREKÖEKÖLET 737
OY 606 TREKÖTMDIÖPMDIÖVELLAKNÖHERÖLMTÖDÖSDÖINRÖSKÖTTADDÖCYVAKÖGSD 665
Db 738 KVÖSNNAED--PKDLM-----LSGERVÖÖTE-----RSVÖSS-----ISLVGÖTD 776
OY 666 Y-----ASSVPTNÖSQÖKSLASÖTÖKELÖGHÜL-----ALÖTÖESHPÖNÖFO 707
Db 777 YGÖESISLÖEVSTLÖGAKÖTPEÖNNCYÖCAAFÖPKÖLIHÖCSÖKDNÖNDPÖBGFTYPLÖGHE 836
OY 708 STÖÖCÖÖTHLÖMEKÖVTIÖAASSPÖLFSHHDÖYIÖE-----AP----- 743
Db 837 VNHRSRTSIEMEE-----SELDAYIANTFKVSKROSFAPSPNGNAEECATF 885
OY 744 TEHWG--RKDAKLTWE-----QFKATRNSPAAT-----CGAOFREGIOAVDL 785

```

```

Db      886 SAHSGSLKOSPVTPECEQKEENOGKNEBNIKPVQTVNITAGFPVVGQDKP-----VDN 941
Qy      786 TSTHWGSSSNVARSQ-----PVAPLDRYAE-RAVNOVYHAR 821
Db      942 AKCSIKGGSFFCLSSQFRGNETGLTPNKHGLQNPYRIPPLPIKSFVTKCKKNLLEE 1001
Qy      822 NFPSITATMEASKLCDRRNAGQVVLVPEKSPMATHLLRMDPSTLASFPNYGTSRNOME 881
Db      1002 NF-----EHSMSPEREMG-----NENIPST-----VSTISRNMR 1032
Qy      882 SOLHNSOYAHN-QYKGSTSTSYGSNLNGKIPLTFE---DLSRHQLDLHRLRPPHRYG 936
Db      1033 ENVFKEASSNINEVGSSTNEVGSSIN-EIGSSDENIQAEELGRNRPKLNAMLR----- 1085
Qy      937 VLGSLQKELANKSENGCTGSGYKGLVSTGTSYHONKREHFE 979
Db      1086 -LG-VLQPEV-----YKOSLPGSNCKHPEIKKOEYE 1114

RESULT 13
US-08-425-061-24
; Sequence 24, Application US/08425061
; Patent No. 5622829
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Clatre
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,061
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1852 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-425-061-24

```

```

Query Match      3.2%; Score 179; DB 1; Length 1852;
Best Local Similarity 17.9%; Pred. No. 6,2e-05;
Matches 222; Conservative 170; Mismatches 381; Indels 470; Gaps 62;
Qy      48 VDBPAQHCHGCHFSIRGVALL--QKKDPKFCSLSRIFHDDKCKDEHKASSPFS--VAKF 103

```

```

Db      31 IKPVPSTKCDHIFCKFCMYLKLNOKKGPSCPLCKNDITKRSLOE-----STRESSQVEEL 86
Qy      104 RRMDCKSLD----- 113
Db      87 LKLTICAFOLDTGLVANSYNFAKENNSPHEHLKDEVSIIOSMGYRRNARRLLQSEBENS 146
Qy      114 -----KLKTSNDGTAPRTLPAKONGT-----SPGCSITFVTSFVPASVGSOK 156
Db      147 LQETSLSVQLSNLGTV-KLTKRKQRIQPOKTSYIELGSDSSBDTYNKATY--CSGIDQE 203
Qy      157 VSPSTOSGKNADRSTLPKSVQEGNDKCNAPSGKNGAAE-ANTDPSMKDLQGPANQD 215
Db      204 LQGIT-----PGTRDEISLSDAKKACEFSETDVTNTEHHQPSNN-- 244
Qy      216 VAANVEDNTSDVYGALPEYPO-----ITWAEVNGADQDPSTKRLSEVVLKRNEDEN 268
Db      245 -----DLNTTTERKRAERPEREKYOGSSVSNLHVEPGGTTHASSLQIENSSLLTKDRM 297
Qy      269 GKTEETLVAEQCNLTGDPNPMGKERDQ-----VAEOCNLTGDPKPVSGOK-----CEQ 317
Db      298 NWEK-----AEFCNKSQOP-----GLANSQHNRMAGSKETCN--DRRTPTTEKKVLDLADP 346
Qy      318 ICNE-----PCEE-----VVLKRSKSKR--KTDKKLKKQKQSKRKRTA 355
Db      347 LCERKEMNKOKLPCSENPREDVEDPVMTLNSIQKVNEMFSRDELLGSDSDHSGSESN 406
Qy      356 ADVSD-----AKLCRRPKKYRLSEITANOVEDSRK 388
Db      407 AKVADVLVDLNEVEDYSGSSEKIDLLASDPHEALIC--KSERVHSKVSERN-IEDKIF 461
Qy      389 DEVHRENAADPCEDDSTIPVPM-----EVSMDIPVSNH-----TVG--EDGL 429
Db      462 GKTYRRKASLP--NLSHVENLILGAFVTEPQIIDERPLTNLKKRRRPTSLGHPDEF 518
Qy      430 KSKNKTFRKSDVVDGSS-----LNMMLN--GKKKRGSVNHYTAHPAGNLSNK 478
Db      519 KKA-DLAVOKTPREINQGTQOTBONGOVNMTTSGHKNKTKGDSIONEKNPNIESLEKE 577
Qy      479 ----KVTPPTASTOHDENDENGDLTNMKTQV-----COHV----- 511
Db      578 SAFKTAEPRISSISNNE-----LELNHNSKAPKKNRLRRKSTPHIALELVYSRNL 631
Qy      512 -----SEISTQRCSSKGTAGLSKGTHTSASTRYGGESTRNQONIHVLSAED-----Q 560
Db      632 SPNCTELOIDSCSS-----SEELKRRKYQMPVRHSRNLQMLGKPKPRGCAKK 660
Qy      561 COMETENSVLSSHAKVSPAHDIOI-----MSDLHE--QSLPRKKKKOKLEY 605
Db      681 SNRPNEQTSKRHSDPTFP--ELKLTNAPSGFTKCSNTSLKEFVNPSLPREKEBEKLET 727
Qy      606 TREKQTMIDIPMDIVELLAKNQHROMLTETDCSDINRIQSTYADDCVIVAAKGDSD 665
Db      738 VKVSNNAED--PKDLM-----LSGERVLQTE-----RSVESSS--ISLVPGTD 776
Qy      666 Y-----ASSVPRTNSQOKSLASQSTOKELQHL-----ALVQESPHPNQ 707
Db      777 YGTQESISLLEVSTLGAKEPKKCVSQCAFEKPGKLHGCSKDNRNDEGFEKRYLGE 836
Qy      708 STQEOOHLRMEEMVTTIAASSPLFSHHDDQYIAE-----AP----- 743
Db      837 VNHSRETSIMEE-----SELDAYLQNTFKVSKROSPAPRSNGNABEBCATF 885
Qy      744 TEHWG--RKDAKLTWE-----QFKATRTNSPAT-----CGAOPRFGIOAVDL 785
Db      886 SAHSGSLKOSPVTPECEQKEENOGKNEBNIKPVQTVNITAGFPVVGQDKP-----VDN 941
Qy      786 TSTHWGSSSNVARSQ-----PVAPLDRYAE-RAVNOVYHAR 821
Db      942 AKCSIKGGSFFCLSSQFRGNETGLTPNKHGLQNPYRIPPLPIKSFVTKCKKNLLEE 1001
Qy      822 NFPSITATMEASKLCDRRNAGQVVLVPEKSPMATHLLRMDPSTLASFPNYGTSRNOME 881

```

Db 1002 NF-----EEHSMSPEREMG-----NENIPST-----VSTISRNIR 1032

QY 882 SOLHNSOYAHN-QYKGSTSTSYGSLNKGKIPLTPE-----DLSRHQLDHLRPLRPHRVG 936

Db 1033 ENVEFEASSSNINEVGSSTINEVGSSIN-EIGSSDENIOAEIGRNRGPKLNMALR-----1085

QY 937 VLSGLQKEIANWSENCGTOSGYKLGSTGTITSHQNRREHE 979

Db 1086 -LG-VLQPEV-----YKQSLPGSNCKHPEIKKOEYE 1114

RESULT 14

US-08-825-886-24

; Sequence 24, Application US/08825886

; Patent No. 5821328

; GENERAL INFORMATION:

; APPLICANT: KING, Mary-Clair

; APPLICANT: FRIEDMAN, Lori

; APPLICANT: OSTERMEYER, Beth

; APPLICANT: ROWELL, Sarah

; APPLICANT: LYNCH, Eric

; APPLICANT: SZABO, Csilla

; APPLICANT: LEE, Ming

; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/825,886

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/425,061

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, Richard A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RA3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 494-8700

; TELEFAX: (415) 494-8771

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 24:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1852 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-825-886-24

Query Match 3.28; Score 179; DB ?; Length 1852;

Best Local Similarity 17.9%; Pred. No. 6.2e-05;

Matches 222; Conservative 170; Mismatches 381; Indels 470; Gaps 62;

QY 48 VDEPAHQCEHFSTRGYVALL--OKDPKFCSLRIFHDYKCKDEHKAASSPFS--VAKF 103

Db 31 IKEVSTKCDHIFCFMCKILNKGKSPQCLCKNDITKSLQE-----STRFSQVLEL 86

QY 104 RRMDCSCLD----- 113

Db 87 LKIIAFOLDGLIEYANSYNFAKKENNSPEHLKDEVISIIQSGVYNNRAKRLILOSEPNPS 146

QY 114 -----KAKTSQNGAPRPLPAKQNGT-----SDGCSITFVASTVPASVSGOK 156

Db 147 LQETSLVQLSNLGTV--TLTKRKQRIQPOKTSVYIELGDSSEEDVYNATY--CSVGQOE 203

QY 157 VSPSTOSSGKNADSTLPKSVQEGNSDKCNAPSGKNGAAE--ANTDSPMKLDQPAQVND 215

Db 204 LLIQIT-----PQSTRREISIDSKAKKAACESEEDVYNTNTHHQPSNN-- 244

QY 216 VAANVEDNTVDYDGLPEVQ-----ITWHTFVNGADQPPSTPKLSEVYLKRNEDN 268

Db 245 -----DLNTEKRAAERHHEKYQSGSVSNLHVPCGNTNTHASSLQIHENSSLLTKDRM 297

QY 269 KTEETLVAEOCNLTKDNPNPSGKERDQ-----VABEOCNLTQDKPKVSGOK-----CEQ 317

Db 298 NVEK-----AEFOCNKSKOP-----GLARSHNNRAGSKETCN--DRPTSTKKYDLNADP 346

QY 318 ICNE-----PCBE-----VVLKRSSKSR-----KTDKLLKKQOHSKRTAQ 355

Db 347 LCEKEMNKOKLPCESENPRDTEDEVWITLNSIQKVNEMFERSDELDGSDSHOGESSEN 406

QY 356 ADVSD-----AKLCRRKPKVLLSTETIANOVEDSR 388

Db 407 AKVADVLDVINEVEYSGSSSEKIDLASDPREALIC-----KSERVHSKVSBN--IEDKIF 461

QY 389 DEVERENADPCEDDRSTIPVPM-----EVSMDIPVSNH-----TVG--EDGL 429

Db 462 GKYTKKRSKP--NLSHVTENLIIIGAVTEPQIOEPRLNKLKRKRPISGLHPEFI 518

QY 430 KSKNKTARKYSDVVDGSS-----LMNLN--GKKRRTGSVHTTVAHPAGNLSNK 478

Db 519 KKA-DLAVQKTPEMINQSTNTEONGQVAMNITNSGHEKRTGDSIONKNPNIESLEKE 577

QY 479 -----KVPRTASTQDDNDQENGLDTNMKTDV-----CQHV----- 511

Db 578 SAFKTAKEPISSISINME-----LELNHNSAKPKKRLRKRKSTRTIHALLVYSNKL 631

QY 512 -----SEISTORCSSKGTAGLSKGTASAFTYGGESTRNGONIHVLSAED-----Q 560

Db 632 SPNPTQELQIDSCS-----SEIKKKYNGQPVHSHNQLQMEKEPATGAKK 680

QY 561 COMETENSVLSHAKVSPAEHDIOI-----MSDIE--OSLPKKKKKQKILEV 605

Db 681 SNKPNEQTSKRHSDTFP--ELKLTNAPGSFTCSNTSELKEFEVNSLPREEKEKLET 737

QY 606 TREKQTMIDIPMIVELLANQHERQIMETDCSDINRISKTTADDQCVIAKQSSD 665

Db 738 VKVSNNAED--PKDLM-----LSGERVLQTE--RSVSSS-----ISLVPGTD 776

QY 666 Y-----ASSVFDTNSOOKSLASQSTQKELOGHL-----ALTTQESPHPQNFQ 707

Db 777 YGTQESISLEVLSTLGAKKATPEPNKCVSCAFENPKGLIHGCSNDNRNDTGFXYPLGHE 836

QY 708 STQEQTHLRMEBVTYLAASSPLFSHDDQYIAE-----AP----- 743

Db 837 VNHSRETSIEME-----SELDAQYLQMTFKVSKRSQSPAPSPNGNAEECATF 885

QY 744 TEHWG--RKQAKKLWE-----QFKATPNSPAAT-----GGAQRRPQIOAVDL 785

Db 886 SAHSGSLKQSPKVTPECEQKEENQGNESNIKPVQYVNTITAGFPVVGQKKP--VDN 941

QY 786 TSTHVMSSSSNVYASRQ-----PVIAPLDRAE--RAVNOYHAR 821

Db 942 AKCSIKGSRRCSSQPRGNTEGLITNNKHGLLONPRIPPLFIKSFVTKCKKNLLE 1001

QY 822 NFPSTIATMEASKLCDRRNAGQVVLYPEKSPATHLRLMMDPSTLASFPNVTGSSRNOME 881

Db 1002 NF-----EEHSMSPEREMG-----NENIPST-----VSTISRNIR 1032

QY 882 SOLHNSOYAHN-QYKGSTSTSYGSLNKGKIPLTPE-----DLSRHQLDHLRPLRPHRVG 936

Db 1033 ENVEFEASSSNINEVGSSTINEVGSSIN-EIGSSDENIOAEIGRNRGPKLNMALR-----1085

OY 937 VLGSLLQKELIANWSENCGTSGYKLGVSCTGTSQHONRKEHFE 979
Db 1086 -LG-VLQPEV-----YKQSLPGSNCKHPEIKKQEYE 1114

RESULT 15

US-08-480-784-2
Sequence 2, Application US/08480784
Patent No. 5693473
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,784
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-784-2

Query Match 3.2%; Score 179; DB 1; Length 1863;
Best Local Similarity 17.9%; Pred. No. 6.2e-05;

Matches 222; Conservative 170; Mismatches 381; Indels 470; Gaps 62;

OY 48 VDEPACHQCHENFSTRGVALL--OKDKPKCSLSRIIFHDKKCODEKHAASSPS--YAKP 103
Db 31 IKEPVSTKCDHIFCKFCMLKLNOKKGPSCPLCKNDITKRSIOE---STRFSOLVEEL 86
OY 104 RRWDCKSLD----- 113
Db 87 LKIIICAFQDLDTGLEIYANSYNPAKKENNSPEHLNDEVSIIOSMGYRNARKLLQSEPNPS 146
OY 114 -----KLKTSNDGTAPRLTPANONGT-----SDGCSITFVSTFPASVGSQK 156
Db 147 LQETSLSVQSLNIGTV-RTLRTRKORIQPKTSYIELGSSSEDTYKATY--CSVGDOE 203
OY 157 VSPSTQSGCKNADRSLTPSVQEGNDKCNAGSKGKGALE-ANTDSPMDLOGPANOYD 215
Db 204 LLOIT-----PGTDEISLSAKKAACEFSETDVTNTEHHQPSNN-- 244
OY 216 VAANVSEDNTSVQALPEVPO-----ITWHIEVGADQPPSTPKLSEVLKRNEDN 268
Db 245 -----DLNTEKRAERHPEKYQSSVSNNLHVEPCGTTHASSLOHENSLLITKDRM 257
OY 269 GKTEETLVAEQCNLTQDPNMSGKERDQ-----VABQCNLTQDKPVSQK-----CEQ 317
Db 298 NVK-----AEFCNKSQKQ-----GLARQHRNRMAGSKETCN--DRPTSTEEKVDLNADP 346
OY 318 ICNE-----PCEE-----VLKRSKSKR-----KIDKMLKQOHSKRTAQ 355
Db 347 LCERKENNKKPLPCSENPRDTEDPWITLNSIQQVNEWFSRDELLGSDSDHGESESN 406
OY 356 ADVSD-----AKLCRRKPKVLLSEIINANOVEDSR 388
Db 407 AKYADVLDVLENDVEYSGSSEKIDLLASDPHEALIC-----KSEYVHKSYESN-IEKIF 461
OY 389 DEVHRENAADPCEDDRSTIPVM-----EVSMDIPVSNH-----TVG---EDGL 429
Db 462 GKTYRKKAFLP--NLSHVENLIIIGAFVTEPQIIQRRPLTNLKKRRRPTSGLHPDEFI 518
OY 430 KSKNKKTKRKYSVYVDGSS-----LMWLN-----GKKKRGSVHHTVAHPAGNSNK 478
Db 519 KKA-DLAVQKTPREMINOGTQTEONGVMMITVNGHENKTKGDSIOEKNNPPIESLEKE 577
OY 479 -----KVTPASTQCHDENDENTENGLDNTMKTVD-----COHV----- 511
Db 578 SAKTKAEPILSSISINNE-----LELNINSAKPKNNRLRRKSTYHIALLEYVSRNL 631
OY 512 -----SEISTQRCSSKGTAGLSKGTSAASTRYGSESTRNGONIHVLSAED-----Q 560
Db 632 SPPNCTELQIDQSS-----SEIIRKKKYNQMPVHSHRNQLMGKKEPATGAKK 680
OY 561 COMETENSVLSHSAKVPAAEHDIQI-----MSDLHE-----QSLRKKKKKKOLEV 605
Db 681 SNKPNEQTSKRHDSDFP--ELKLTNAPGSPFKCSNTSELKEFVNSLPREEKEELEF 737
OY 606 TREKOTMIDIDPMDIVELAKNQHROLMETDCSDINRLOSQTADDCVIYAAKGS 665
Db 738 VKVSNNAED--PKDLM-----LSEGRVQTE-----RSVSSS-----ISLVGCTD 776
OY 666 Y-----ASSVFPTNSQOKSLASOSTOKELOGLH-----ALTOESPHPNQF 707
Db 777 YGTQESISLLEVSTLTKAKTEPNKCVSCAAFEPRKGLIHGCSKDNNDREGFVPLGHE 836
OY 708 STQEOQTHLMEEMVTTAAASSPLFSHHDDOYIAE-----AP----- 743
Db 837 VNHSRETSIEME--SELDADQYLQTFPVSKRSQSFAPSNQMAEBECATF 885
OY 744 TEHWG--RKDAKLLTWE-----QFKATRNSPAAT-----CGAOFPRGICQAVDL 785
Db 886 SAHSGSLAKOSPKVTFPECEKEENQGNKESIKIKVQIVINITAGFPVVGQDKP----VDN 941
OY 786 TSTHVMGSSSNVYASRQ-----PVIALDRLYAE-RAVNOVHAR 821
Db 942 AKCSIKGSRCLSSQFRGNETGLITPNKHGLLONPYRIPPLPFIKSFVYTKCKKNLLE 1001

```
OY      822 N$P$T$A$T$M$E$K$S$C$I$D$R$N$G$Q$V$U$L$P$K$M$P$A$T$H$L$R$M$D$S$T$A$P$N$T$G$S$R$O$M$E      861
        ||::|||
Db      1002 N$-----E$H$S$P$E$R$E$M$G$-----N$EN$B$P$T$-----V$S$T$R$N$N$R      1032
        ||::|||
OY      882 S$Q$L$N$S$O$Y$A$N$H-$Q$Y$K$G$S$T$Y$S$G$N$K$G$P$L$T$F$E$----D$L$R$H$D$H$D$L$H$P$R$H$P$R$V$G      936
        ||::|||
Db      1033 E$N$V$K$E$K$E$S$S$N$I$N$E$V$G$S$T$E$V$G$S$S$T$N-$E$L$G$S$D$E$N$I$Q$E$L$B$R$N$G$P$X$L$M$L$R$-----      1085
        ||::|||
OY      937 V$L$G$S$L$O$K$E$I$A$N$S$E$N$G$T$O$S$G$Y$K$L$G$V$S$T$G$T$S$H$O$M$N$R$E$H$P$E      979
        ||::|||
Db      1086 -L$G-$V$L$O$P$E$V$-----Y$K$Q$L$P$S$N$C$K$H$P$E$I$K$Q$E$Y$E      1114
```

Search completed: September 16, 2002, 22:24:02
Job time: 7446 sec


```

Db 342 TPEKGRNRREFOVYDEVPVSLPCEISO--EGIKEDHAIIPSK--RSTPAHSILFTGNDVSP 397
QY 409 VPMVEYSMDIPVSNHTVGEGLKSSKNKTKRKRYSDVDDCSSLM-----WMLNKK--KRIG 462
Db 398 CP-----PGQRT--ERKLSLPKKTAK--PYIDCKKSTVIFSMSIDSSOVNSHG 445
QY 463 SVHHTVAHPAGNLNKKVYPTASTOHDENDTENGIDPTIMHTD--VQHVSEISTORCS 520
Db 446 PSMNTVSQTRDLNKGKV-----GGLFDIRLASDQYFKRYSQVNDKPTT 490
QY 521 S-----KGTAGLSKGTHTSASTRKYGESTRNGCNHIVLSAEDOCOMETENSVLS 571
Db 491 SLIHODNDVYSRDAEPCLRDIFSSSSKSSSGGWLRTG--VDIYDFRNHNNTNNSSEPSN 548
QY 572 HSAKVPAPHEIDIOIMSD-----HEKSLPCKKKKKOLEVTRERKQIMI 613
Db 549 LKIRPPESTEVADLSKVLQKADASADRKCTVMQEHGAPRSOSHDKRETTTEONN- 607
QY 614 DDIPIVDIELLAKNOHERQLM--TETDCSDINRIOSKTTTDDCVIAAKGSDVYASVFD 672
Db 608 DDIPIVELLAKNOYERCLPDREEDVSNKQPSQETAKR SKNALLIDINETYDNGISIED 667
QY 673 TNSOO--KSLASOSTQKLOGLHALITQESPH-----PON--FOSTOEOQTHLR 717
Db 668 NNTSRPPPCSSNAKREE--HFRPGROQNSHDEFPISC PYVSPFGIIFPTQENR--- 720
QY 718 MEEMVTIASSPLFSHDDQYIAEAPTEHWGRKDAKLTWEOFKATINSAPAATGA--- 774
Db 721 -----ASSIRFSGHNCMLGNLPY--GNQNPSPSEFVYLA-----CDTQOSPVN 764
QY 775 QFPGIOAVDLTSHVWGSSSNVYASROPVIALPDYAEFAVNOYHARNPSTIATMEASK 834
Db 765 QYR-----EASHPIWPS-----MIPPOQYKPVSNINOSTN--PGTIS--QASN 806
QY 835 LCDRRNAGVYLYKE-----SMPTILLEM 861
Db 807 NENWNLNFAVANGKCKGCPPEPFCGCKHAGVSSSSRPIDNFSSSSIPALHLSIL 866
QY 862 DPSTLASPP--NYGTS--SRNOMESQILNSQYAHNQYKSTSTSGYMLNCKIPTFEDL 917
Db 867 DPLRSTPADQHGNTKFTKRHPRANGSKFEIQTGDSKSAVSTK--QIP--FLY 921
QY 918 SRHOLHDLHPLRP--HPRVYVGLSLOKEILANMSENGCTSGYKLVSTGITSQOMNRE 976
Db 922 SKRPTQESRRSPFITPPIGR--SLSFONASWSPH-----HOEKTK 962
QY 977 HEFALNMGFSAKWNALOLGVSASD-----PLSARNSIAQSWTRGKRVHPL----- 1026
Db 963 RKDI-----FAPVYNTHEKVPYFASSNDQAKROLGASNS-----NMLPLKFEHNT 1006
QY 1027 DREYRQD-----ICITNKNPADFTTISNDNEYM 1054
Db 1007 DKEKKQKRAKESCNNNASAGPVKNSSGPVICSVNRNPADFTIPEPGNYM 1056

RESULT 2
B47328
natural killer cell tumor-recognition protein - mouse
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tum
C:Species: Mus musculus (house mouse)
C:Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C:Accession: B47328; I77662
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cells
A:Reference number: A47328; MUID:93133824
A:Accession: B47328
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1507 <AND>
A:Cross-references: GB:L04289; NID:9192866
A>Note: authors translated the codon AGT for residue 972 as Arg
R:Infret, A.; Anderson, S.K.

```

```

MOL: Immunol. 30, 1307-1313, 1993
A:Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing
A:Reference number: 157820; MUID:94019422
A:Accession: I77662
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 235-237; 263-294 <RIN>
A:Cross-references: GB:S65998; NID:9425701; PIDN:AA828500.1; PID:9425702
C:Genetics:
A:Gene: NK-TR
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

Query Match 4.1%; Score 225; DB 2; Length 1507;
Best Local Similarity 19.5%; Pred. No. 0.00031;
Matches 239; Conservative 169; Mismatches 459; Indels 356; Gaps 52;

QY 93 ASSPFSVAKPRRMCSCCLKLTKTSQNTAPRTLPKQNGTSDGCTTFVASTVPASV 152
Db 214 AASRPY--ADVAVIDCGVLAITL--TKDVFEKKRKKPTCSEG-----SD 253
QY 153 GSQKVPSTOSSQGNADSTL-----PK-----SVQGNDSKNAPSGK 192
Db 254 SSSRSSSSSESESEVERETIRRRHRRKRVRAKRRKREMSSEPRKRTVSPG 313
QY 193 NGAAEANTDSPKDLQGAONTAVANVSEDNTSVGALPEVPO-----TWH 241
Db 314 SERSDVN-----EKRSVDSNTRKREKPVYRPEIIPVPRNRLRDMPAITVE 361
QY 242 IEVN-----GADQPS-----TPLSLEVLYKRNDENGKTEETILVA 277
Db 362 PEONTPDPAVYSDQKPEVSKSGKRIKRGITRIYTHTPRSHSHSKDDSETPPHKE 421
QY 278 EOCNLT--KDNPMGKERDQVAEOCNLTDPKPYSGQKCEQICNEPEEVLKRSKSK 335
Db 422 EMORLAVRPPSGEKWSGDKLSDPCSSRMDBERSLSQSRMSVGYSDLSSTARHSGH 481
QY 336 RKTDKKLKMQOHSKKRQAQADVSDAKLR--RKPKYRL-----SEITANOV 383
Db 482 HKHNRK--EKKKHKKKAKQ-----KHCRHRRQTKRRRIYMPDLERSPTHRKSSCV 534
QY 384 EDSR--SDVHRENADPCEDRSTIPVPMVEYSMDIPVSNHTVGEGLKSSK----- 433
Db 535 RERSRASSSSSHHSSKRDWMSKQDDGASATHSSRDSYRSKSHSRSSRSRRSAVSK 594
QY 434 -----NKKRRYS-----DYYDGSLSM----- 451
Db 595 SSSRLNRSKSSRSSRSRGRRTSISPKKQALSENKPVKTEPLRPVPOGCVLQPVAA 654
QY 452 -----NMLNGKK-----KRTGSVHHTVAHPAGNLNKKVYPTAS 485
Db 655 ENIVYIPLSDSPPSRMPKPGOKPMKPSYERIOEMKAKTTHLLPVQSTYSLTNIRATVSSS 714
QY 486 TQHDENDTENGIDPTNMHTDVC-----QHVSEISTORCSKGTAGLSK----- 530
Db 715 SYHKREKPEESDGSAYKSYSDRSSGSSGSKSSRSRSRSTRSRSLRSLR 774
QY 531 ---EKTHSASTKYGGESTRNGCNHIVLSAEDOCOMETEN--SVLSHS----- 573
Db 775 SPSSRSHPNKYSDGSHSSRSSTSVSSDDGRAMFRSNKKSVTSHKRRHSNSEKTLH 834
QY 574 -----AKVSPAEHDIOIMSD-----LHEO--SLPCKKKKKOLEVTRERKQIMI 615
Db 835 SKYVRGEKSSRRHKKYSBSRSSSLDYTSDSDSHQVYSAPEKEQGVLEALNDQ---- 889
QY 616 IPMDIVELLAKNOHERQLMTEPCSDINRIOSKTTTADDCVIAAKGSDVYASVPTNS 675
Db 890 -----GKGREGKPKPEWECPRSKKESEEDHSDSV-----SKGNKNGASKMPSSES 936
QY 676 QOKSLASOSTQKLOGLHALITQESPHPONFOSTOEOQTHLRMEEMVTIASSPLFSHHD 735

```

Db 937 NSEODVTSKRKSDPRRG---SEKEGEASSDSESEVGOSHIIK-----AKP----- 979
Qy 736 DOYTAETAEIEM--GRKAKKILTWEOFKATITRNSPAATCGAOFRCIGAVDLTSTHWGCS 793
Db 980 ---AKPTSTFTPLPGSDGAMKSRPOSSASESSCSNIG--NIRGEPOKOKHSDKDLKGD 1034
Qy 794 SSNVSROPVLAFLDRAVERAIVQVHANRPSTIATMEASKICDRNMGQVLYLPKESMP 853
Db 1035 HTKRA-REKSKAKKDK-KHKAPKQAFHWOPPLFEGDDE--ELMNGKQVYQDKREKRH 1090
Qy 854 AT-----HLLRMWDPTSLASFPNYGTSSRNOMESOLH--NSOAHNQYKGSTST 900
Db 1091 VSEKCEAVKDGIDIPNVEKCDDEGSSPKPKGTLEQDPLAEGHOPSSCPALPKVEDNTAS 1150
Qy 901 S-----YGSNNGKIPPLTFED-----LSRQQLDHLRLPLRP 931
Db 1151 SPSSAQLHEEHGPG-GGEDVYDQTDNDMEICPPDRTSPAKGEVYSLANRILDSPEVNIIP 1209
Qy 932 -----HPRVGLSLLOKEITANMSENGCYKGLGVSTGTS--HQMNRKEHEALN 982
Db 1210 EODECMAPRAC--GE--QESMSSESKILG--ESGVKODSSTVTSPEVETSGKKEGAEKQ 1264
Qy 983 SGMSAKNNALQ-LGSVS-----SSADFLSARNSIAQSWTRG----KGK-MVHPLDRF 1029
Db 1265 MWL-TDKMKPLQGVGNLSVSTATSSALDYKALSTVEVKPQGLRIEIKSKHKVGRGSLF 1323
Qy 1030 --VRQDICTIKKNPADFTTISND 1050
Db 1324 DEVRTARL-NRRPRNOESSDD 1345

RESULT 3
T00385
KIAA0624 protein - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00385
R:Shikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880
A:Accession: T00385
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1983 <ISH>
A:Cross-references: EMBL:AB014524; NID:g3327061; PIDN:BA01599.1; PID:g3327062
C:Genetics:
A:Note: KIAA0624

Query Match 4.0%; Score 221; DB 2; Length 1983;
Best Local Similarity 18.6%; Pred. No. 0.00074;
Matches 210; Conservative 171; Mismatches 423; Indels 328; Gaps 53;

Qy 83 HDO-----KKCODEKASSPSPVAKFRMDCSKLDIKTSDNCTAPRTLPKAKNGTSDGCS 139
Db 947 HDEVDVAKCHSPPRNERGKGIKIRH--ISCTEKLKTESISVP-----TSDHRS 995
Qy 140 ITFVSTFVPASVGSOKYS-----PSTQSS--OGKNADRSTLPKSVQEG----- 181
Db 996 L-----IEANQNSKVSLEDITICTLPKRSSFLIHRQSGSKIMASLNGPPPPPI 1048
Qy 182 -----NDSKCNAPSGKNGCAEANTDSPKDIQGAQNYDVANVSEDNTSYD 228
Db 1049 KNNVEDAMGNVYMLNKFSPSPESANECSKVLSDSA--LEAFEAETFRMTNVKSSGCTSVR 1105
Qy 229 VGALP-----EVQITWHIE-----VNGADQPPSPKLESEVYIKKNEDENGKT 271
Db 1106 KGPLPPLINRAMSCSPGEPHASTGREGRKKPLTSGMDASELTFRAWERITISPEVSDS-- 1163
Qy 272 EETLVAEQCNLTR-----DPNP-----MSGKERDQ 296
Db 1164 -----VRQDSLTKRQHKENKQFQYTEKEBKMAASRSRYALNSEDPLPCSDLSGKERK 1218

Qy 297 VAEOCNLTGDKPRPVSGQKCEQICNEPCEVYVLKRSSKRRKTDKMLMKKQSHSK----- 351
Db 1219 TLHKVK-TTSTFVSAGDE---DNVCKLEVASIYTLPRKPKKRCNLQDYTOTNTLLI 1273
Qy 352 RTQAQADVSAKLCKRRPKK-----VRLSEIINANQVEDSFSEDEYHRENAADP 399
Db 1274 ESPQVETETFPNALEKDKQNTSTREGSGTPSCENLKMVNSQOTLTTEMTAFRLSNRP 1333
Qy 400 CHDD-----RSTIPRVEMSDIPVSNHTVGEEDGLKSSKKTKYKRYSDVYDDGSSLN 452
Db 1334 LAPTLQEMASVAAVASLPEESKAREIFSDNLAKTPLGDSENKKER----- 1379
Qy 453 WLNKKKRTGSHVHYAHAGNLNKKYPTASTQDDENDENGDTVMHKTQVQJHS 512
Db 1380 ---GKLLQSETLHTSLMLORKNVSEK--SENCOQSISSNSGCPSLPA-LSEVINISQOT 1434
Qy 513 EISTORCSSKGT-----AGLSKGTHTSAASTKYG--GESTRNGONIHVLSAEDOCOMET 565
Db 1435 RRSMECTSGRAIPPTGSKCPQKHDTSTAVGDSSGSPREGRG---DIGTNGQKMT 1490
Qy 566 ENSVLSHSAK---VSPAHDIQMSDLHE-----OSLPK--KKQKLEVTREKOT 611
Db 1491 -NKTLSHSEQYFALTPALHKIQLGEEIOSDEPNLESLOSREPRLPQROEAMNTEESRK- 1548
Qy 612 MIDDIPMDIVELLAKNQHEQRLMTEDCSDINRIQSKTTAADCVYVAKKDSY----- 666
Db 1549 -----AEDEMKSAKDOPSLPEGNK--NKTNLD--LVKGNRSVYKHRLA 1590
Qy 667 ---ASSVEDTN--SQOKSLAS---OSTOKELQHLALTTOE-----SPHPQFOS--- 708
Db 1591 AMSKASRKFPADVSPRRVATITFPQSGSRSGDHLSTGVECNPLFPPEPTPSASIDE 1650
Qy 709 --TQEOQTHLMRE-----MTTAA SPLPSHHDOY--IAEAPTEHMGK 749
Db 1651 SRLSENRRKHVKKSENLPTFVLPNREPSTHVSQKNSNLSORQNEFKKVVSESPKHEMS 1710
Qy 750 KDAKKTWQFATITRNSPAATCGAOFRCIGAVDLTSTHWGSSSNVYASROVIAPLR 809
Db 1711 KD-----VTAQNLVRESGAP-----SPITTSLEAEFSSNQRLSPFP-PLER 1754
Qy 810 YERAVNQVHARNFSTIATMEASKICDRNMGQVLYLPKESMPATHLIRMDPSTLASF 869
Db 1755 -AQK-----SRVSSPLASFLQOQRSASL-----EMWPEPHLYR---SKLSKSI 1794
Qy 870 PNYGTSSRNOMESQLHNSQYAHNQYKSTSTGSLNGLKIPLTPEBDLSRHQDLHLRLP 929
Db 1795 NVHGDLRLKS---HPVYREHNFSESTSI--DNALSRLTLGNEFSVYNGYSRRFRGF 1846
Qy 930 RHPFRVGLGSLLOKEITANMSENGCYKGLGVSTGITSQNMNRKEHEALNSGMFSK 989
Db 1847 SELPSCD--GN-----ESMAVRSYGTGTGPRSAISI-----PRPIDYGLFGKE 1886
Qy 990 WNALQGSVSSADFLSARNSIAQSWTRGKGVHPLDRFRODICTYKNP 1041
Db 1887 QQLAFLENVKRS--LTQGRLMKPSFLKNPG-----FLKDL-----RNP 1923

RESULT 4
562982
vacuolar protein VAC7 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein N2467; protein YNL054w
C:Species: Saccharomyces cerevisiae
C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 07-May-1999
C:Accession: S62982; S58722; S68168
R:Berez, P.; Dolignon, F.; Crouzet, M.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62975
A:Accession: S62982
A:Molecule type: DNA
A:Residues: 1-1165 <BER>
A:Cross-references: EMBL:Z71330; NID:g1301918; PID:e339893; PID:g1301921; MIPS:YNL054
A:Experimental source: strain S288C

R:Berger, P.; Dolignon, F.; Crouzet, M.
 Yeast 11, 967-974, 1995
 A:Title: The sequence of a 44 420 bp fragment located on the left arm of chromosome XIV
 A:Reference number: S58711; MUID:96021608
 A:Accession: S58722
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1001 <BEW>
 A:Cross-references: EMBL:U12141
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
 R:Berger, P.; Dolignon, F.; Crouzet, M.
 Yeast 12, 297, 1996
 A:Title: Corrigendum to: the sequence of a 44 458 bp fragment located on the left arm of
 A:Reference number: S68168; MUID:97060022
 A:Accession: S68168
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1165 <BEF>
 A:Cross-references: EMBL:U12141; MID:91314216; PID:91098487
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
 C:Genetics:
 A:Gene: SGD:VAC7
 A:Cross-references: SGD:S0004999; MIPS:YML054w
 A:Map position: 14L
 C:Function:
 A:Description: required for normal vacuole inheritance and morphology
 C:Keywords: transmembrane protein; yeast vacuole
 F:925-941/Domain: transmembrane #status predicted <TM>

Query Match 3.9%; Score 217; DB 2; Length 1165;
 Best Local Similarity 18.7%; Pred. No. 0.0003;
 Matches 204; Conservative 156; Mismatches 416; Indels 314; Gaps 46;

QY 155 QKVPSTOSQGNKADSTLPKSVQEGNDSKCNAPSGKN;AAEANTDSPM--KDLQGPQA 212
 Db 6 RKLVEIEYEAAPANNLL-----SNNSVVAPNPST;SASTST-SPLREIYDDSYA 58
 QY 213 NYDVANVSE-----DNTSVYDVALPEVPOITWHIEVNAADQPPSTPKLSEVYLKRNED 266
 Db 59 TANTTSNVYOHNLPTIINMLDSDATSH-NQDHWHSIDIN;AGTSMST---SDIPTDLHLE 114
 QY 267 ENGCTEETIVAEQCNLTKDPRPMGKERDQVAEQCNLTK;PKPVSGQKCEQICNEPCEV 326
 Db 115 HIGSVSST-NNNNNALINHPISLH-----STIPVMEVSMDFVSNHVTGED 139
 QY 327 VLKRSKSKRTDKLMMKQOHSKRTAQADYSPAKLCR;KPKVYRLSEITINANOVEDS 386
 Db 140 -LSMPSSSLNKKSSLLVANS-----PAPASVDELK;KPAVISNNMPTSNIALYOTA 191
 QY 387 RSDVEHRENAADPCEDR-----STIPVMEVSMDFVSNHVTGED 427
 Db 192 RSAIHHPSSSTASAKAFKASAFENNTAPSTNIGSNT;PAPL-----LPLPS----- 240
 QY 428 GLSKSKKTKT--RKYSDYVDGSSLM--NWLNGKKRT;SVHHTVAHPAGNLSNKKY-T 481
 Db 241 --ISQOKKPKTIEPTMHVNSREILL;GENLDDTKAKN;PANSST-THDNGVPANDGLRI 297
 QY 482 PTASTDHDDENGDLDTNMH-----KTDVCOHVSEIS;QORCS-----KGTGA 526
 Db 298 PNHNNADNENNNMKKNKKNINSKNERNDTSKICTT;FTAPASTAPLGSTDTMTQALTA 357
 QY 527 GLSKGKTHASASTYKGSESTRNGONIHVLSAEDCCQMET;ENSVLSHSAKVPAPHDIOIM 586
 Db 358 SVSSSNADNHNHNNKKTSSNNNGNNSASAKTNAIDIKN;NADLASTSNNAIND---- 413
 QY 587 SDLHEGSLPKKKKKQ-----KLEVTREKQTMID;ILPMDIVELL-----AKNQ 628
 Db 414 -DSHESSEKPYTADFAARLATAVGENEISDSEETFYVS;AANSTNKLJFPDSSSOQOQ 472
 QY 629 HEROLMETDCSDINRLOSQTTA---DDGCVIYAKGDSYAS-----SVFEDNSQ 676
 Db 473 QQQQPPKQQQQQQQNGHTISATISAPLLNNKKLSRLKNSNHISTGAILNMTATISTPN 532

QY 677 QKSLASOSTOKELOCH-----LALTQESPHPQNPQSTQEQOQTHLMEMWTIASSPLF 731
 Db 533 LNSNVQNNNNIMSGHNHLDLSLTKQEPH-----QLQQQPPMDVOSVSYSDNP-- 585
 QY 732 SHHDQYIAEAPTEHMGKDAKKLTFEQFKATRTNSPATC-----GAOPRPGIOAVDL 785
 Db 586 ---DSNVIAKSPDKRSSLVLSKSPHLLSSSSNGNTISCPNVATNSQLEPN---NDI 639
 QY 786 TSTHYMGSS-----SNVASHQ-----PVIAPLDHYA-----ER 813
 Db 640 STKKSLSNSTRHSSANNSNVGDKRPLRTTVSKIFPSNPNGAPLRRYSQVDPHVNLED 699
 QY 814 AVNQYHANPSTIATMNSKLCODRNAGQVYLYKESMPATHLLRMDDPSTIASFPYKG 873
 Db 700 YIEQPH--NYPLMNSVAKDEFFNSRN-----NKP--HGINTFYGDNNVLEENNG 746
 QY 874 TSS-----RNMESQLHNSQYAH----- 891
 Db 747 DSNVNRQHTNLQHEFIPEDNESDENDIHSMEFYNNHKNDELTKPLISDYGEDDYDDYD 806
 QY 892 -----NOYKSGTSTSYGSNINNGKIPL-----TFEDL-----SRHQLDHLRPLRPHR 934
 Db 807 RPNATFNSYGSASNTHELPLHGRMPSRNDYDFVGNNTGNNOLE--YTPLRMRKG 865
 QY 935 VGVLGSLQKEIANNSEKCG-----TOSGYKLGVSITGTSHOW--NRKEHFPALNSGMFSA 988
 Db 866 QRHL-SRTNNNSIMNGSIHMGNDQVTHSINNNDYGVSPINFTSRKSPFYKVK----- 918
 QY 989 KWNALQGSVSSS-----ADFLSARNSIAQSWTRGKGKVVHPLDRVRO-----DIC 1035
 Db 919 --NFLYLAFLVLSILTMGFLIGFLATNKELODVD-----VVWMDNVISSDLELIFDIT 970
 QY 1036 ITNNKPADFT 1045
 Db 971 VSAFNPGRFS 980

RESULT 5
 T34513
 hypothetical protein ZK783.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34513
 R:Favella, A.; Vaudin, M.
 submitted to the EMBL data library, August 1994
 A:Description: The sequence of C. elegans cosmid ZK783.
 A:Reference number: Z21536
 A:Accession: T34513
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3507 <PAV>
 A:Cross-references: EMBL:U13646; PIDN:AA24418.1; GSPDB:GM00021; CESP:ZK783.1
 A:Experimental source: strain Bristol N2; clone ZK783
 C:Genetics:
 A:Gene: CESP:ZK783.1
 A:Map position: 3
 A:Insertions: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 3.9%; Score 215.5; DB 2; Length 3507;
 Best Local Similarity 17.9%; Pred. No. 0.0031;
 Matches 242; Conservative 179; Mismatches 510; Indels 419; Gaps 52;

QY 12 RYVGTNCMLAGTGAVAPVLELTATPRODAA-----EAGVDEPAQHC- 56
 Db 1780 RMGSGCGCKMAGYGDATCIKIEEPKSDKTACTDMSRLCELEKKQCYVDEBEVQCG 1839
 QY 57 -----EHFSIRGVALLQ-----KKDPKCSLSRIF----- 82
 Db 1840 ACLPFGHHPINGTCSLSLISGLCAQKNDCKNKAECIDIHDPHSFCSCPDGFIGDGMICDDV 1899

```
QY 83 -----HDOKCKEHAHSSPFSYAKFRMDCKCL-----DLK----- 116
D 1900 DECANAGCDEKTCENTISFNCVCEGFKYD-EKCVUDEKOPAREKIEIDENSS 1958
QY 117 -----TSNGAPRPLPAKONGTSDGCSITVR---STFVASVSGKVS 158
D 1959 SSNGOEKPTTKGIVSSSTAASSSTTAEPHTTISSTSTKMTSSKSEPMVTMSSES 2018
QY 159 P-TOSSOGKANADRSITLPKSVQENDSKCANAPSGKNGAANANTDSPKMDQGPANQNDV 216
D 2019 PEVSTSSSKTTASSETTVSSSTPSSSSSEAPLTSSPATTEVITSSYKS----- 2068
QY 217 AANVEDNTSVQALPEVPOITWHIEVNGADOPSTPKLSEVVLKRNEDENGKTEETLV 276
D 2069 TTPKESSSEITVLKSSKSPVET---ESSYKSSPSTPTSSQSVTSVTPETSKS--IVL 2122
QY 277 AEQCNLTK-----DPPMSKERDQVAEQCNL-----TKDKP----- 308
D 2123 SSEAPVTSPTPEVHTSSETPKSLASSTTGDPTNSTPTSSLASVKSSTSAPEGTSASVA 2182
QY 309 PVSGOKCQOICNEP-----CEEVYLKRSSKSKRTDKMLKKQOHSKKTAAQADVSDA 361
D 2183 PVKLSLSPDVSOPTKTFDAETESTVQASSETSGTSYKSTSEPSHVKLTSTSSNPS 2242
QY 362 KLCRRKPKKVKL-----SEIINANO---VEDSRSDVHREN 395
D 2243 SVPTSPKSTPTVESTEQPTSTPSSGSLTPMNSSEVLTTSEPHVLSLSLSPVSGSS 2302
QY 396 AADCEDDRSTIPVPMESMDIPVSNH-----TVGEDGLKSSKKTTRKYSDDVVD 446
D 2303 TTPNNLSSESTVEP-KTSEVSLNSEPSTTEAPTLSPILSTTNNLSQSSSTVED 2361
QY 447 GSSLNMLNGKKKRTGS---VHHTVAPRAGNLKKTPTASTQIDD--ENDTEGLDTN 501
D 2362 RSEIIS--ENSEKPTSAELVTSYTHVA---SSPDVPTSSSEDDLTGSTER----- 2411
QY 502 MHKTIVCOHSEISTORCSKGTAGLSKGTSHASTKYGEGESTRNOQNIHVLSAEDQC 561
D 2412 -----IPEASSKQIISSTPTPTDTTASEEPTKTSMPD----- 2445
QY 562 QMETENSVLSH-----AKVSPAHDIQIMSDLHESLPPKKKKOKLEVTREKOTM 612
D 2446 -LSTTANVLSSESTPSSSKSPVSSSTEGISVYTSSTFESKVPSESTISSVEEDLTKT- 2503
QY 613 IDDIPIMDVELLAKNOHERQMTEDCSDIRIOKTTADDCVYAAKSDSVASVFD 672
D 2504 ---PSPILEETTASETSEPLTEDSLVSVRIHELTSSN---VPKSESTYTTSSSS 2556
QY 673 TNSOO-----KSLASQSTQKEL-----OGHLALT-----QESPAPQ 704
D 2557 KPQGEPPAGILSTVYVPTSSVSLTASIEAITSNTPPKQGRPTTTPPKSLVKTSTSPS 2616
QY 705 NFQSTOEQOHLRMEENVTIAASSPLFSHDDOYIAEAPTEHMGKRAKLTWEOFK-- 761
D 2617 TVTSESESESTKRTVSTTVSTTT-----PTEETTTSESLITLAAPSKPTE 2662
QY 762 ATTNSPAATCGAO-----FRPGIOAVDLTSTHWS-----SSNYSARQVI 804
D 2663 STSSSEAPTPAKTSETKPKSNVSTSKSTEVE-ISTSSGSLSESTSTSESEPTE 2721
QY 805 APLDRAERAVNOVHAR-----FPTSTATMEASKLCR----- 838
D 2722 APATVSEASSTLLEENSSPTSSSEASVKSLSLFPESI-TSNAVYSSAPAEITMS 2780
QY 839 -----RMAQVLYLPRESKPAHLLRMDPS--TLASFPNYGSSSNQMESOLHNSQY-- 889
D 2781 SESHREISTVSESEPE--PEIPLSTVSPNVVTAASSIP-----SEEPILSVTSSSTPRV 2833
QY 890 -----AH-NQYKSTSTSYGSLNGKILPLFELSHNQ----- 921
D 2834 RLITGTPDDLIVSVTVPSHGNKRONITASSVPSNSTSPILIPSESLTTPQPPPTTTAK 2893
QY 922 -----LHDLHRPLRPHPRGVGLSLQKEIAMNSENGCTSGYKLGAVST 965
```

```
D 2894 PATTSKRGPSIOPPAEMFTTPAPPPNSNGYGETNOEEOVYSTTTTEAP-----SLCS 2950
QY 966 GITSQMKRKEHFEALNSGMSAKMNAQLG-----SVSSADFLS-----AANSIAQSWT 1016
D 2951 TVTCHSLATCEB-----STGVICIRDRGFIDGDTTACSSKSTADICISLPSLCADKACDNST 3006
QY 1017 R-----GKGKMY--HPLDRFVRDQC 1035
D 3007 RSCBCDAGTIGDYVCSPPHQCVCVLRDMLC 3036
```

```
RESULT 6
microtubule binding protein D-CLIP-190 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13030
R:Plantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A:Title: A class VI unconventional myosin is associated with a homologue of a microtu
A:Reference number: 217588; MUID:9613549
A:Accession: T13030
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-1690 <LAN>
A:Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AA896783.1
A:Experimental source: strain Oregon R
C:Genetics:
A:Cross-references: FlyBase:FBgn0020503
C:Keywords: cytoskeleton
```

```
Query Match 3.9%; Score 215; DB 2; Length 1690;
Best Local Similarity 19.1%; Pred. No. 0.0013;
Matches 171; Conservative 139; Mismatches 352; Indels 234; Gaps 33;
QY 174 LPSVOE-----GNDSKCNAPSCKKGAEMANTDSPM---KDLGPAONYDVAA 218
D 401 LOKNINELKARIVELESALGNERK-----KTEELQCSIDAEQFGDELANOASOYKKEI 454
QY 219 NVSEDNSTVDGALPEVPOITWHIEVNGADOPSTPKLSEVVLKRNEDENGKTEETLVAE 278
D 455 HDLESKITIKLVSAITPSLSIL-----PPDLP-----SDGALQBEIQL 493
QY 279 QCNLTKDPNPSGKERDQVAEQCNLTQDKPKVYSGKCEQIC-----NEPCEVYVK 329
D 494 QEKWTIOQKEVESRIAEOLLEEORLRENVKYLN---EQIATLOSELVSKDEALEKFSLS 549
QY 330 RSSKSKRTDKKMLKKQOHSKKTRTAOADVSDAKLCRRKPKKVKRLSEIINANOVEDSRSD 389
D 550 ECGIENLRLEALLKEENKQAEQAEFT--RKLKSEVLEVLSSSELONLKATSDLSLES 608
QY 390 EVHRENAADPCEDDSTTPVPMESVM-DIPVSNHTVGEDGLSKSNKTKRKYSDDVDDGS 448
D 609 E--RVNKSDECE-----ILQTEVMRDEQIELNOQJDEVTQJLVQKA-----DSS 653
QY 449 SLAMNLRNKKKRTGVSVHTVAPRAGNLKKTPTASTOHHDENDENGLOTNMKTQVC 508
D 654 ALDDMLRLQKEGT-----EEKSTLLEKTEKELVQIKEQAAKTLQKEOLE 698
QY 509 QHVSEISTORCSKGTAGLSKGTSHSAS--TKYGEGESTRNOQNIHVLSAEQOCOMETEN 567
D 699 KOISDLQOLAEQK-----LVAREKTEMAINOJLEKESIEOALAKONLEPFOKKOSFS 753
QY 568 SVLSHSAVSPAEDHDIOM-----SDLHESLPPKKKKOKL--EYTRKQOTMIDI 616
D 754 EYHLEIETAKNOHQKDELLEVESGESLKLQOQLEKTLGHEKILQALAELEKKEKETIEXK 813
QY 617 PMDIYELLAKNOHERQMTEDCSDIRIOKTTTADDCVYAAADGSSYASSVFDTSQ 676
D 814 EOELOOIOSKSASESALKVQOV-QLEOLOQOAAA-----SGEGSKTVAKLHDEISQ 865
```

Db 941 ECTELVKSPPKKMKGVDE--KL-ERTPEKKIAESTTPPAKKIKLNRTGKKIGAENAS 997

QY 398 --DECEDRSTPIRPMESMDIPVSNHTGEGEDGSSKNKTKRKYSVVDGSIAMML 454

Db 998 TTKRESEKLESTSSKIKQEKVAKKRVASGESS--TVDYTSTSSGSPVAKS 1053

QY 455 NGKKRTGSGVHTVAH-----PAGNLGKKVTFPTASTOHDDENDTENGIDTNMHTDV 507

Db 1054 EKKTTTTRTVIKTMEEYNNNDTAPAEVDYIMIHVQSKWMDDEFESEEDVKT----- 1106

QY 508 CQHVSEISQRCSSKGTAGLSKG-KTHSASTKYGKSGSTNGQIINHLSMDDCQMETE 566

Db 1107 -----TQPIQSGKSSIIKNVTKPSAKTKESEQPEKIQKL-----PKFAS 1152

QY 567 NSVLSH---SAKVSPEAHDIQIMSDLHEOSLPKKKKKQLEVTREKQTMIDDPMDIYEL 623

Db 1153 HELMGLHELRSSKGSASSSEKRAKREHSGSEKDNDRKKSQAQPDKESTVD-----R 1204

QY 624 LAKNOHEQMLTETDCSDINRIQSKTTADDDCVYAA-KDSQDVASSVFTNQQOKSIAS 682

Db 1205 LSEQGHFHTLSSQSK-----ETRTSEKHSHSVRSSNKKDPTPGKKQKVDYDSRD---YS 1254

QY 683 OSTOKELQGHALATLQESPHPONFQSTQEOOTHLMEEMVYTLAASPLFSH----- 733

Db 1255 SKRDEKGEELA-RKKDSD-PRGKESLSGQSKLEEDLPRKGAESKSSNSPPRODKP 1312

QY 734 --HDDQYIAEAPTEHMGKRDAKKLTW---EQFKATRNSPAATGCAQDPGIOAVDLTST 788

Db 1313 HDHKAPYETFKRCEE--TKRPDKISGREREKHAARNA----- 1348

QY 789 HVMGSSSNYASQPIAIPDLRYAER--AVNOYHARNFSTIATMFAKLCD--RRNAGOV 844

Db 1349 --GDSGSGGKLPCILNPPDLPMKEELAVGOVEKSAVVKPQLSHSSLSDLTRETDEA 1405

QY 845 VLYP-----KESMPATHLLRMMPDST-----LASFP---NYGSS 876

Db 1406 AEFDPYNSDESQSVSYKEEBAVASISKDLKEKTEKKESITVATPASQPCADNSQSQSS 1465

QY 877 RNMQESQHLNSQYAHNOYKGSTSYSGSNLNGKITPLTFEEDLSRQHLHRLPLRP 932

Db 1466 PVSFSPRSHSPGSGQSTRSHSSASAGSQ-DSKKKKKKKKKKKKKKKKKKKH 1520

RESULT 8

T06310

hypothetical protein Fl1C18.80 - Arabidopsis thaliana

C1Species: Arabidopsis thaliana (mouse-ear cress)

C1Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

C1Accession: T06310

R1:Bevan, M.; Terry, N.; Argiles, W.; Buysbaert, C.; Dasseville, R.; De Clerck, R.; ewes, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, April 1999

A1:Reference number: Z15569

A1Accession: T06310

A1Molecule type: DNA

A1Residues: 1-852 <BEV>

A1Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:Fl1C18.80

A1Experimental source: cultivar Columbia; BAC clone Fl1C18

C1Genetics:

A1Gene: ATSP:Fl1C18.80

A1Map position: 4

A1Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3

Query Match 3.8%; Score 209.5; DB 2; Length 852;

Best Local Similarity 21.2%; Pred. No. 0.001;

Matches 174; Conservative 97; Mismatches 282; Indels 269; Gaps 39;

QY 31 VLELTAIPRQ-----DAAAEAGVDEPNOGCEHFSIRGY-----VALLOKK 71

Db 129 ILETVAKYRSCVMDLECDALLIEMFOHLK--AIRDHSGNVFSSKEMNTITVLESE 186

0Y	72	DPKCSLSRIFHDQK	KDDEHKASSPFSVAK	FRMDCSCDLKLT	-----	SDNGTAPR	125		
Db	187	DIPSEMLSPILHSHV	KDKDEDSIOVS	-----	RRLAEVOLSCASCKALXTYL	TEAVKSSGV	238		
0Y	126	TLPKAKNGTSDGCS	ITFVRSTFVPASVSGOKVS	PSYSSOGSKNADRS	TLPKSVQGNDSK	185			
Db	239	PLDKYSINIVASIG	EGT	-----	SALQODVYANEKEDSOGHI	KRETEYERKAELSTPER	292		
0Y	186	CNAP	---SGK	-----	NGAAEAN	-----	TDSPMK	229	
Db	293	TDAPKDESGKSGV	NGVAQNDSSVPTDSK	KDDOTGANGDEPQJ	LDNPRNTDLNNT	----	349		
0Y	230	GALPEVQIITHT	-----	YNGAD	-----	OPSTPKLS	257		
Db	350	---BEKPDVEHQH	IEKENESSVKOADLSK	ODIKETEPEALDLSK	DYLTSP	PPDVSSVT	406		
0Y	258	EVLKRN	-----	DEGNKTEETLVA	BECONLT	KDPMNPMASGERRQV	AEQNL	303	
Db	407	AATSENENKKSUO	ILPDKTSGDETANVSS	PSAEELPQOSV	PKKTANOKKE	-----	SS	461	
0Y	304	TKDRPVSGOKCE	ICNEP	---CEEVYLKRS	SKRRTDKLKLK	KOOSHKKRPAQADVSDA	361		
Db	462	TEEVKPSASIA	TEEVSEPTSEPOV	TKKSGK	KVASSSKTPV	TPVPSKST	-----	SET	515
0Y	362	KLCRRKRRKRL	SEIITNANOVEDSND	VEYHRENNAAD	PCEDDRSTI	PVPMVEYSMDI	PVSN	421	
Db	516	KVAQOSEKVV	-----	YSDNAOESTKPR	KEKKKPRGKA	IDESL	-----	555	
0Y	422	HTVGED	-----	GLKSSKN	-----	TKRKS	-----	DYVDGSSL	450
Db	556	HTSSGDNKPAV	SSGKLASSKKEAKOY	EEPSNSNTKRRSL	GGOGKASGESLV	---GSRI	613		
0Y	451	MNV	-----	LANGKKRTGVS	HHHTVAHPAGN	-----	LSNKVVTPASTQ	---HD	489
Db	614	KVMPMPDQAYY	KGVVESYDAKKK	-----	HLVIYDGDG	EILYLNKKMSP	PLDEBELSD	668	
0Y	490	DENDTENGCLT	NMAKTDVCOHVEIS	TORCSSKGTAGL	SKGKTHS	-----	AASTKYGE	544	
Db	669	EEAADOTGOE	-----	EDASTVSGAGS	KAKKATTPASKS	SKTSODDKT	PASKSKDKE	719	
0Y	545	STRAGONIHLV	ASBDOCMETENS	SVLSHSAKVS	PARHDIOIMSDL	HEQSLPRKKKKOLE	604		
Db	720	ASREE	-----	ASSEEEBEEB	PPRYGKSGS	---SRKKD	ISSVSKSGSK	SAASSKKKEEESK	774
0Y	605	VT	---REKQIMID	IPMDIYELLAKNO	HEROLMTETDC	SINRIOSKTTAD	DDCVIAAK	661	
Db	775	ATTSKSKSGPV	KSV	-----	AASKTGK	-----	KAK	801	
0Y	662	DGSDYASV	PTPTNSOOKSLAS	OSTOKELQCHL	ALTTQES	PHP	703		
Db	802	SGS	---AST	---PASKAKES	SESESE	-----	TPKPEP	830	

RESULT 9
T24587
hypothetical protein T06E4.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24587
R:Lloyd, C.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19910
A:Accession: T24587
A:Status: preliminary; translated from GB/EMBL/DBAJ
A:Molecule type: DNA
A:Residues: 1-1295 <NTL>
A:Cross-references: EMBL:Z70756; PIDN:CAA94789.1; GSPDB:GN00023; CESP:T06E4.1.1
A:Experimental source: clone T06E4
C:Genetics:
A:Gene: CESP:T06E4.1
A:Map position: 5
A:Introns: 17/3; 218/2; 385/3; 481/3; 946/3; 1034/3; 1222/2

Query Match	3 8%	Score 208.5	DB 2	Length 1295
Best Local Similarity	18.4%	Fred. No. 0.002		
Matches 208	Conservative 179	Mismatches 402	Indels 339	Gaps 52
QY	84 DKKCKDEHKAASSPSPVAKFRMRDCKSLDKLT--SDNGTARTPLPAKONGTSDCSITNF	142		
DB	96 DMSVCEPPASSA-----HVDPNLEQTKYVSEDSISKDCSTADSFATDG-DVTL	145		
QY	143 VNSTVPASVSGQKVPSPSTOSSGKNADRSSTLPKSVQEGNDSCNAPSGKNG---AAEA	198		
DB	146 ANDFQOATIQOLENTPST-----PLRSQTVGDESTPAPSLRSKIPRIESEV	194		
QY	199 NNDSPMKDLOGPAON-----YD---VAANVEDNTSVQ---VGALEPVQITWHI	242		
DB	195 VOSTVESVOSTSQHEKSIETSLYEFEMLENNIEHDEYVKDLQITELGKHRTLV	253		
QY	243 EVNGA---DQPSYTKLSEVVLK-RNEDENGTEETLVAE---QCNLTKDPNPMSCKE	293		
DB	254 EIKERREVEKMLMOQVEEAIVAKTKSEBAATLEQLLEKRIELEMKLT-BPN---	308		
QY	294 RQOVAEQCLTDPKPVSOQKEQJCNCECEVYLKRSK--SKRTDVKLKLKQOHSK	351		
DB	309 KLOFEHQLEELK-----SRCEELTKALKVDVMQHSIEDYEKKFVELQEKKEEDEL	361		
QY	352 PTAQADVSPAKLCRRPKPKVRLLS-----EIIANQVQEDRS-----DEVHR-EVA	396		
DB	362 QAKAKEDIETLQ-----KVELETITNKREVPNSLETETLKSHEIYKRLMLDEIRHLENG	416		
QY	397 ADPCEDDRSTIEV-PWEYSMD--IPVSNHTVGEDGLKSS---KNKTKRKYSDVVDGSS	448		
DB	417 MSALQPKNDTELEELQKTLIDDLKIDCCMLTQSKMLETQSELVEVEKATSEIGEAQKNG	476		
QY	449 SIAMNMLNGKKKRTGSV-----HHYAHAGNLSNKKVPTASTQND-----D	490		
DB	477 ELLEQNTSLRVENAKLVDMEGOLDNHHKRAEDKVRISETLTTTIESLRDQSEASDKLMD	536		
QY	491 ENDTEGLDITNNKHKIDVQCNHVEISITORSSKGTAGLSKGTHTSAATKYGGESTRNQO	550		
DB	537 SSTQNEYSALENT-----VSELETMRREYKA-----	564		
QY	551 NIHVLASBQCCOMETENSVLSSHAKVSPAEHDIQIMS-DLHQSPLPKKKKKQ-KLEVTR	607		
DB	565 -----SYDVVCSIQLELEIEIOHETSVLEAEIRIKLELDAQEAAYKSSQDLKKEIYO	619		
QY	608 EK-QTMIDIDIPMDIYELL-----AKNOHERQ-----LMTETCCSDIN--	643		
DB	620 EDCQKRLQDLKEQIQOQLVSLRETSVMHQESARHQEKKYQIQSKLMSFEAYEILRSSI	679		
QY	644 -----RIQSTTADDCVIT-----VAAKDGSDYASS---VFDTNSQOKSL-	680		
DB	680 DSIQAEVVRQOSADQDKHILIEDYLKIRQAEETNKLKSLDIASSQEDILDKNOESLIT	739		
QY	681 -----ASQSTQKOLECHLALTQOESPHQNFQSTORQOOTHILRMEVNTTAAASPFS	732		
DB	740 DDLKELHSAESTNQOLEQVLSLEMLKIEVSNAR--QKVMSVYLKESEFALQLELSA--S	794		
QY	733 HHDDQYIAAPTPEHMGKDKAKLTLWEQFATRTNRSPAATCGAIFRGIAVDLTSTHWG	792		
DB	795 QEVSRSYVDAAYQ--EKDGLRLVDTLKLKIEDTEKS-----AODLOOSSV--	838		
QY	793 SSSNNTASRPVIAPIEDRAVNOYHANFSTIATMTASKLCIDRRNAGOVVLYLKESM	852		
DB	839 -----BEIKOLOLDLQNFQMAEVLDSL-----NEKL	865		
QY	853 PATHLRMDPSTLASFPYQGTSSRNQMSQQLHNSQYANQYKGSYFGSNLKNKIPIL	912		
DB	866 NSH--KRDWALAS-----QLEELQHKLVVGSQYEVNKEELIGAKI-----	906		
QY	913 FREDLSRHOLDLHRLPHRPNVGLVGLSLQKEIANSWENCGTOSGKVLGVTGTISHOM	972		
DB	907 ---NKKKEVDELN-----AKGIDLE-----GME---ELKKSILEVSEAVV	940		

QY 973 NRKEH-----FEALNSGFSAKMMALDGVSSVSSADFLSARNISIAQ 1013
Db 941 QRRREELIAQVSKHRDQEOLOLTLDELKSAOSHSTETSRSGNELAARLEELASISPAQ 1000
QY 1014 SMTREKGMVHPLDFVRDODICITKNPA-----DFTT--ISNNEY 1053
Db 1001 KALQVEDVKH-----OODIQISEANEAMVKLKODFETERTSIONEF 1042

RESULT 10
A53577
ascites statloglycoprotein 1 - rat (fragments)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 12-Apr-1995 #sequence.revision 12-Apr-1995 #text.change 07-Feb-1997
C:Accession: A53577
R:Wu, K.; Fregien, N.; Cariaway, K.L.
J. Biol. Chem. 269, 11950-11955, 1994
A:Title: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifurcated
A:Reference number: A53577; MUID:94216302
A:Accession: A53577
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1630 <MUA>
A:Cross-references: CB:U06746
C:Keywords: glycoprotein

Query Match 3.8%; Score 208; DB 2; Length 1630;
Best Local Similarity 20.2%; Pred. No. 0.0023;
Matches 206; Conservative 120; Mismatches 104; Indels 292; Gaps 41;

QY 106 WDSCKLDKLTSDNGAPRLPAKQNGTSDGSIYFVR3TFVPASVSGQKSPSTOS-S 164
Db 13 WLCSLCL-----CSCLLLPVNTSTPSAKPTSTALP3STNPSQMTSQQVSNPTASSYR 64
QY 165 QGRNADR-----STLPKSVQEGNDKCNAPSGKN3AAEANT----- 200
Db 65 MTKMTGASPMVTSSTITTLPOSGHTGSMKTRNPO--TTTTEVTTLTSSASSDQVYET 123
QY 201 -----DSPKMDLQPAONVDVANYSEDNTYVDVGLAPEV-----PQITWH 241
Db 124 TSQTTLSPTDTTSHAPRESSPSTSVITLTSTSTEST3SDTGHMAVTTQGSTPATTE 183
QY 242 IEVNGADQPPSTPKLSEV-----VLKRNEDENGTEETL---V 276
Db 184 ISVT-----PSTQKMSPTSTSTQTEITTLSSQHTGCKATTNNPQTGTTEVTTTISA 238
QY 277 AEQGNLTKDPRPMGSKERDOVAEOCNLT-----KDXKPVV---GQKCEQICNEPCE 324
Db 239 SSSDHPSTSSPSTPGNTAPRTTETSTTTTKVLMTSLOQKLPSTGLSTQTELTTLPEQS 298
QY 325 E--VVKRSSKSKRRRTKILM-----KROHSHKRTAQADVSDAKLCRRKPKV 371
Db 299 QHTGIMKTTSTQTTTLEVTYTRTLSSSDHRAETSSJYTTLEPDTTTSHPRESSPP 358
QY 372 RLSEITINANOVEDSRSDVH---RENADPCEDDSTIPVPEVSMIDPV----- 419
Db 359 S-TSVILTHGREGTSGDTGHMAVTTQGSTPATTEISVTPSTQKMS---PVSTFSTQ 414
QY 420 -----SNHTVGEDGKSSKNKTKKRYSDV---VDDGSLMAMWLKNGKRRKTSVHHTV 468
Db 415 EITTLSSQHT---GGMKTYTRNPORTTTEVTYTLSSASISDQVQVETTSATLSPDTTT 471
QY 469 AHPAGNLNKKVTP-----TASTQHDNDENTJNGDLTNMHHKTPVQHVSEIS 515
Db 472 TSNHPSVSSSSPSPSTEGTSVDGTGLTAVTTQDSTPATVQGLSTSSQSTLSTVSPSTLS 531
QY 516 TORCSSGKTAGLSKGTTHSAASTKYGESTRNGCNIHVA--EDQCCOME-TENSVLSH 572
Db 532 TQETSTDELTS--SOSQHTGSMKTTNHPQTTRNTEVTTTJSSASSDQVQVETTSQTLSD 589
QY 573 SAKVSPAEHDIQIMSDLHEQSLPKKKKKKLEVTRKQTHIIDLPIMIVELLAKNQHRO 632

Db 590 ATTTSHAP-----RESSSPSTST-----DILTTMAS----- 615
QY 633 LMTFDCDINRIKQKTAADDCVYAAKDGSDVASS-VFQTNQOK-----SLASQSTOK 687
Db 616 --TEGTSQDTGHMAVTT-----QGSTPATTEISVTPSTQKMSPVSTFSTQTE 662
QY 688 ELQGLHALTQESPPHPONFQSTQEOQTHLRMEEMWTIAASSPFLSHDQYIAE----- 741
Db 663 -----ITTLSSQHTGSMKTTNHPQTGTTEVTTLSASS-----SDQVQAEISSQTT 710
QY 742 -----APTEHMGKDAKKLWEQFKATTRNSPATCGAOPRFGIADVLSTHWGS--- 793
Db 711 LSPDTTTSHPRESSSPSTSDMLTTT---ASTEGTSCDPTG---HTTAVTTQGSIPA 762
QY 794 -----SSNVAHQRPVIAPIADRRAERAVNQHARNFSTIATMEASGLCRNRNAGVYLTPK 849
Db 763 TQOLSTTASQ-----KMSVTSPTTSTQELSTLPQSOHTGSMET--- 803
QY 850 ESMPTHLRLMMDPSTLASFPVYGTSSRNQMESQLHNSQVYAHNOYKSGSTTS-----YG 903
Db 804 SSRPQTTSV-----TSTLSSPSGSTRPV-----QTRSVTSSDEKTNPTS 843
QY 904 SNLNGKITLFEEDLSRHQLHDLRPLRPHRVGVLSLQKEIANWSENCGTQSGYKLGV 963
Db 844 SGVSTNSPATTEVILPTSPSE-STPGNTAPRTTETSTTTTKVLMTSLOQKLPSTGLTGT 902
QY 964 ST 965
Db 903 ST 904

RESULT 11
T44231
hypotheical protein U86 [imported] - human herpesvirus 6 (strain 229)
C:Species: human herpesvirus 6
A:Variatey: strain 229
C:Date: 21-Jan-2000 #sequence.revision 21-Jan-2000 #text.change 02-Jun-2000
C:Accession: T44231
R:Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A:Title: Human herpesvirus 6b genome sequence: coding content and comparison with hum
A:Reference number: 222734; MUID:99412318
A:Accession: T44231
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1520 <DOM>
A:Cross-references: EMBL:AF157706; PIDN:AA049674.1
A:Experimental source: strain 229; variant B
C:Genetics:
A:Note: U86

Query Match 3.8%; Score 207.5; DB 2; Length 1520;
Best Local Similarity 18.6%; Pred. No. 0.0028;
Matches 200; Conservative 170; Mismatches 480; Indels 227; Gaps 41;

QY 84 DOKCDEHKASSPSPYAKFRPMDCSKLDKL-KTSDNGTAPRLTPAK---QNGTSDGCS 139
Db 270 DNKEHKQOYSTPDALISLR-PSSQKMDLKKITTYEVLP--LPKALIDNGSYCYVD 326
QY 140 ITFVRSFVPASVSGQVSPSTQSGKRNADRTLPK-SVQEGNDKCNAPSGKNGAAEA 198
Db 327 TOKYKTHIRK-----SIQTKGHS--QTIHKKVQKNHEN--HVS----- 366
QY 199 NTDSPPMDLQPAONVYAAVNSDNTSV--VQALBEVQITTHIEVNGADQPPSTPKL 256
Db 367 --RSDLKQKRS-NQHEDEAVTEARDFSKLDPLSLPLMTPE-----PTL 407
QY 257 SEVYVK---RNEDNGKTEETLVAEQCNLTRKDRPMGSKERDOVAEOCNLT---DPKPV 310
Db 408 NFAVHKTKIHSDSELTHTTKKIHRSKTSL--QDRVLSKNAAPRAPYTDNKKYKHHDKDT 465
QY 311 SGKCEQICNEPCEBEVVLKSSKSKRRKTDKLMKKQHSKRRTAQADVSDAKLCRRKPK 370

```

Db      466  NPK-----MKHSRGRTSKKNKNSNGHDVNEAVKNSGKAATSK 511
Qy      371  VLLSEIINANOVEDSRDEVRHENAADPCEDRSTIPVPMESMDIPVSNHTVGEDLK 430
Db      512  ----SD--PSHMLHKTSDEQYKTS-----PDNEKISTPPKSKTHCHDSSSEGGY 559
Qy      431  SKKNTKRYSDVVDGSSLLMMMLNKKKRTGSVHTTAHPGN-LSNK-----VVTPT 483
Db      560  KSPNSESNT-----GNCUSDKFYDNFTTTPN 587
Qy      484  ASTOHDEDDTENG-----LDTNMRKTDV-----COHVS 512
Db      588  SKTNKTESTENTDLNSNNTNTLETEDSNIIOPFQLCNETIITSTACPTQETP 647
Qy      513  EISTORCSSKGTAGLSKG-----KTHS-----ASTKYGESTRNGONIHVLSAD 559
Db      648  STMNRCAKSKHAGANKMLTDNSPIRSHSNPSSFTAFKNSNGNSTMS-----TSNGD 702
Qy      560  OCOMENENVLSHSAKVSPAEDIOIMSDLHOSLPKKKKKOLEVTRREKOTMIDDIPMD 619
Db      703  ECTDKKPN-----CSSTENKSETSNOTNGENSDKPLSKTFTTEVSDRASSRASS 754
Qy      620  IVELLAKNOHEROLMTETDCSDINRIQSK--TTADDCYIVAAKDGSDYASVFTDINSQ 677
Db      755  RASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASS 814
Qy      678  KSLASOSTKEIQLGHALITTOESPHPONOSTO-EDQTHLRMEVNTIAASPLFHHDD 736
Db      815  RASGRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASS 874
Qy      737  QYIAEAPTEHMGKQKQKLTWEQFATTRNSPATGAGFRPGIOAVDLSTHVMSSSN 796
Db      875  RASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASSRASS 934
Qy      797  YASROPVIALPLRYAERAVNOVHARNEPSTIATMEA-SKLCDRRNAGOVV-----LYP 848
Db      935  RASGRASSRASSRASSRASSRAGKTPNNKLMISNPSPQTEKTSNKRKRQIYCDNKROIY 994
Qy      849  KESMPATHLRMADPSTLASFPNYGTSRNQMSQLHNSQYAHNOY-----KSTST 900
Db      995  HQTISTEAVESEIKRCPPELIFYKNAARLQSFNNHNDQFYRPRPHIRTRKRKSEST 1054
Qy      901  SVGSNLNGKIPLETFEDLSRHQLDHLRPLRPHRVGLSLQKELANSENCGTOSGYK 960
Db      1055  NITDSESS--TSKRSRSHKSPDLSLTPKRRKKSQ--SSSISSTI--EENRSMSPRE 1107
Qy      961  LGVSTGIT--SHQNRKKEHFEALNSGMFSAKWNAQLQGSVSSAD--FLSARNSIAQ-- 1013
Db      1108  TGTDKLITFNYQHSRTSSSSSSSVSSSSSSSKSRKSKYKNSKEDPATYSPSRLOQLE 1167
Qy      1014  -----SMTRGKGMVHP-----LDRFVRODICTITKNKPADFTTISDNENYMDY 1056
Db      1168  KOTRPSASASRDKTKIKSPNHESKRRHADMF-RNSQKTGEKFLPDLNPSPTHTHOSNH 1223

RESULT  12
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagianakis, G.; Siden-Klamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:

```

```

A:Cross-references: FlyBase:FBgn025392
A:introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match          3.7%; Score 204.5; DB 2; Length 5327;
Best Local Similarity 19.1%; Pred. No. 0.021;
Matches 225; Conservative 174; Mismatches 457; Indels 319; Gaps 51;

Qy      31  VLELTATPRQDAAGVDEPAHQCE--HFSIRGYVA---LQKDKPFCSLSRITH 83
Db      1967  VAEKSPLEKSEASRPASVAESIKDEAEKSEESRESVAEKSPLEKASRPASVAESIK 2026
Qy      84  DQ-KKCDEN-----KASSPSPVAKFRMDCSKLDLTKTSNDNTARPTLP 128
Db      2027  DEAEKSEESRESVAEKSPLEKASRPASVAESIKDEAEKSEESR-RESAAEKSPLE 2085
Qy      129  AKONGTSDGCSITFVSTFVPASVG--SQKVPSTQSSQGNADRSITLPKSVQEGNDSK 185
Db      2086  SKRASR-----PASYAESVDEADKSKSEESRESMAESCKNAQSI-KGDQSP 2130
Qy      186  CNAPSCKNGAENANTDSPMKDLQGPANQDVAAVSEDNTPVDYALFEVQITWHI EVN 245
Db      2131  LKEVSRPEESVAKDDPVKSKKE-PSRRESVAGSVTADSRADQSPLE-----ESK 2179
Qy      246  GADQPPS-TPKLESVVLKRNEDENGTEETLVADQCNTLTKDPRPMGSKERDQVADQCNTL 304
Db      2180  GASRPESVVDVSVDEAEKQSRRESSTESYI-----PPAKDDKSKPEVLQPVSWT 2230
Qy      305  KDKRPVSGQCEQICNEPCEVVLKRSKSKRKRTDKMLKQOHSKRTQAQVDSAPKLC 364
Db      2231  ETRREDADQPMKRSQAEKRESRESTIAESTKASSPDEKSLASKASRCSVAESI---KYD 2287
Qy      365  RRPKPKVRLSEITIANQVEDSRSDVEYHRENAADPCEDRSTIPVPMESMDIPV-SNHT 423
Db      2288  LDKPQIITK-----DDKSTEHRSRESL-----EDKSAVTSSEKSVSRPLSVASDHE 2331
Qy      424  VG---EDGLKSS--KNKTKRKY-----SPVVDGCS--SLMMMLNG 456
Db      2332  AAVAIEDDAKSSIPKDKSPGFAEVVSSPIEBATMEFSKIEVEKSSIALSLQSGSG 2391
Qy      457  KKK-----RTGSVHTVAAHPAGNLNKKVTPTAS-----TQHDDENDTENGDLTNM 503
Db      2392  KQDTDSPPVAVAGDPSHAVA-----SVSTVPTLTKPALAQIGAAKTVSSPLDEALR 2445
Qy      504  KTDVCOHVS-----EISTORCSSKGTAGLSGKTH-----SASTKYKGESTRN 548
Db      2446  TPSAPEHISRADSPAECASEIASODKSPQVLEKSSRPANVAESKDDAOLKSSVEDLRS 2505
Qy      549  GQNIHVLSAE---DQCOMTEFNSVLSHS---AKVSPAER-----DIQIMSDLHESLP- 595
Db      2506  ---PVASTEISRPAASGETASSPIEBAPDPFAEFQEAELVPLTLELGNLPTLSSPV 2561
Qy      596  -----KKKKKOLEVTRREKOTMIDDIPMDIYELLAKNOHEROLMTETDCS-----D 641
Db      2562  DVAAHSAVQPAELSKVDLEKRASSPIDEAPKSLIGSPAEREPESAESAKAABAEVEKSKD 2621
Qy      642  INR---IQSKTTAD-----DD----- 654
Db      2622  ASRPSPSVESTKADSTKGDISTSPESVLEGPDKDVEKSESRSPSVASITGDSTKQVS 2681
Qy      655  ---CVIYAAMDGSDYA-----SSVYDT--NSQOKSLASOSTKQELQGHALITQEE 699
Db      2682  RPASVYESVDEHDKAESRESIAKVESVIDEAGKSDSKS--SSQDSQDKDKSTLA-SKEA 2739
Qy      700  SPHPNQSTQEOQTHLRMEVNTIAASPLFHHDDQYIAEAPTEHMGKRAKTLWEQ 759
Db      2740  SRRESVYESKDKDAEKSESPESVIASGEV-----PREKSLDSDKDTST--- 2784
Qy      760  FKATTRNSPATGAGFRPGIOAVDLSTHVMSSSNVYASROPVIALPLRYAERAVNOVH 819
Db      2785  -----RPSVYESVYAE--DEKSEQQRRESVAVSVAKDTFKDKQSQ 2824

```

[illegible]

RESULT 13

Hypothetical protein F12F3.3 - *Caenorhabditis elegans*
 C.Species: *Caenorhabditis elegans*
 C.Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C.Accession: J134418
 R.Fulton, B.: Wohldmann, P.
 submitted to the EMBL Data Library, July 1998
 A.Description: The sequence of C. elegans cosmid #12F3.
 A.Reference number: 221521
 A.Accession: J134418
 A.Status: preliminary; translated from GB/EMBL/DDJ
 A.Molecule type: DNA
 A.Residues: 1-1488 <Full>
 A.Cross-references: EMBL:U08002; PDB:1AC25865.1; GSPDB:GN00023; CESP.F12F3.3
 A.Experimental source: Strain Bristol N2; Clone F12F3
 C.Genetics:
 A.Gene: CESP.F12F3.3
 A.Map position: 5
 A.Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match	3.7%	Score 204	DB 3	Length 3468
Best Local Similarity	18.5%	Pred. NO. 0.013		
Matches 214	Conservative 161	Mismatches 119	Indels 364	Gaps 46

[illegible]

Dd	769	PAESEÖKTAEVNAKAKÖKEVDNLRKREAEVAAKKTADEKLTLEAEANIKKTALEVAEK	828
Qy	364	CRRPKRVLLSEIIANOVED-----SKSDEVHRE--NAADPCEDDRSTI	407
Dd	829	ÖKEDEBÖLKTLEFVESAKSAAEKTLEKQAOIKKAAEADAVKKÖKELNEKNKLEAAKSA	888
Qy	408	PVPMESMDIPVSNHVTVEGDLKSSANK-----TKRYSVDVDDGSSLM	452
Dd	889	ADKLKEESAASKSVSESVFGEKKTKAGEKTVÖVESEPTSKTIDTKVGAETPA	948
Qy	453	VLNGKK-----KRTGSVHNHTVAPRAGLNNKK-----VTEPASTQHD	489
Dd	949	DEFPKKIITIKKTEKSDSSIÖKSALDSEVKÖKÖDEDEPTRAVSETOVTEADSKQ	1008
Qy	490	DENDTENGIDTNM-----HKTDVCÖHVESEISTÖRCSKGTAGLSKGTSHSAS	538
Dd	1009	KETDEKLTLDIAETAKTKÖADEKSKLDQÖKTKVSEDDAARKEKLN-DKLTLESEIA	1067
Qy	539	TKYGG-----ESTNGÖNIHYLSA-----EDOCOMETENSVLSISAVSPREHIOI	565
Dd	1068	TKASADKLTLEBÖAOAKKAAEVAEAKKÖKEKDÖLKTLEAASKAAAELELEKQAOI	1127
Qy	586	MSDLHÖSLIPKK---KKÖKLEVTR-----KÖTMIDIDIPMDIVELLAK	626
Dd	1128	KKAGADVAKKÖKEDEKKNLEANKKSAGKLTIEBSAAKSQTYEÖAKLD-AÖYKAK	1186
Qy	627	NOHERÖMETTDCSDINRIÖSKTTADD-----DCVYAAK	661
Dd	1187	TA-EKÖTKLEKESKTESSEKETEVDKPKKVKLKKTEKSDSSIÖKSETSKTVESAG	1245
Qy	662	DGSDVASYEDTNSÖK-----SIASÖSTÖK	687
Dd	1246	PSESEÖKYAADAAKÖKEDEKÖKLEAEITAKKSADKSLKLEASKLKAAEVAEAKKÖK	1305
Qy	688	ELÖGHLALTÖÖSPHÖNFÖSTÖ-EOÖTHLMEEWVTIAASSPLFSHNDÖYTAEDPTEH	746
Dd	1306	EKEÖÖKLTLETAASKRAAAEKTLEKÖSHIKKAAEVD-AVKKÖKELEKÖRLESEATK-	1363
Qy	747	WGRKÖDKKLTWÖFKFNTNRSPAAÖG-----AÖFRPGIOAVDLTGRHNWSSS	795
Dd	1364	--KADKEKLEBÖK-----KKAETALLETÖKEÖKLAÖÖSRLEDEAKKSÖKÖKLES	1416
Qy	796	NVASÖPVIAPLDRVYAEAVNOY-----HANFPSTI---ATMEAS-KLDC	837
Dd	1417	ETRSKÖTEAPKRSVDEKPKKVKLKKTEKSDSDSIÖKSASASTYDAATLSDFLVE	1476
Qy	838	RNRAGOVVLTPEKSPMAT	855
Dd	1477	KKTVOKEVÖSPDESTAT	1494

RESULT 14

hypothetical protein SA2447 [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, M. A.; Mizutani-Hiti, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Seki, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
L:ancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KUR>
A:Cross-References: GB:BA000018; PID:g13702612; PIDN:BA843752.1; GSPDB:GND0149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447.

Query Match 3.7%; Score 202.5; DB 2; Length 2271;
Best Local Similarity 18.3%; Pred. No. 0.0088;
Matches 152; Conservative 138; Mismatches 415; Indels 259; Gaps 34;

QY 90 EHRASSPFVAFRMDCSKL-DKLTSDNGTAPRTLPKONGTSDGSIYFVSTFV 148
DB 1025 DSKSMVSSSMSTSGSSTSESLSDSGSTSDSKSLSTSGSGSTSTSTSTST 1077
QY 149 PASV---GSGKVBPSTSGSGKKNADSTLPKSVQEGNDKCNAPSGKGAEAANTDPMK 205
DB 1078 SASVRTSESOSTSGSMASGSDSMSTSTSTSTSTSTSTSTSTSTSTSTST 1131
QY 206 DLGPPQNVYVAAVNSVDNTPGALPEVPOQTWHEVNGADCPSTPRLSEVFLRNE 265
DB 1132 STSGSVST 1174
QY 266 DENGKTEETLVACNLTQKPNPMSKERDQVACNLTQKPNPMSQKCEQICNEPCEE 325
DB 1175 TSDISEAISEST 1234
QY 326 VVLRSSKSKRTDKTKMKQOHSKRTAQADVSDAKLCRRK----- 367
DB 1235 TSLSDSTSESGSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1294
QY 368 -----PKVRLSEIINANOVEDSRDEYH-----RENAADCEDDRS 405
DB 1295 TSLSTSLSDSTSDSKSLSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1354
QY 406 TIFPVMESMDIPVSNHTVGEDGLKSSKNKTKRYSDVDGSLMNLWNGKKRTGSVH 465
DB 1355 KSD-----STSMISMSGSTG-----STSTSTSTSTSTSTSTSTSTST 1402
QY 466 HTVAHPGANSNKKVPTATOHDENDENGDLTNMHTKDVQCHVET-STORCSSKK 524
DB 1403 NSASOSASN-----STSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1447
QY 525 TAGLSK-----GKTHSASTRKYGESRNGONTHVLSAEQOC-----METENS 568
DB 1448 ST 1506
QY 569 -----VLSHSAKVPAPEDHDIQMSDLHEOSLPKKKKKQKLEVTREKOTMIDIPMDI 623
DB 1507 ASST 1553
QY 624 LAKNOHERQMLTETDCSDIRIOSTKTTADDCVIAAKGSDVYASVFDINSOOKSLA-S 682
DB 1554 -----SLSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1598
QY 683 QSTOKELQHLALTTOESPHPONQSTOEOQTHLMEMEVTIASSPLF-SHHDDQYIA 740
DB 1599 QSTSR-----STASGSASSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1642
QY 741 EAPTEHWRKDAKLTWEQKATTRNSPAACTGCAQFRPGIQAADLTSTHWGSSSNYASR 800
DB 1643 LST 1695
QY 801 QPVIAPLDYRAERAVNOHARNPEPTIATMEASKLCDRRNAGQVLYPKESPATHLIRM 860
DB 1696 -----VMSAS-1SDSQSMSESV----- 1711
QY 861 MDPSTLASFPNYGSSRNOMESOLAHNOYKKGSTSTSYGNLNGKIDPLTFEDLSRH 920
DB 1712 -----NDSSESSES-----NSESQSKMSGSTSVSDGSL----- 1742
QY 921 QLDHLRPLRPHRPGVGLSLQKEIRANMSENGCTOGYKGVSTGTSTHOMNRKEHEFA 980
DB 1743 -----VSTSLKSESVSSSLSGSQ--SMDSVSTSDSSSLSTSLRSE-SVSESDSL 1795
QY 981 LNSGMSAKNNALQLGVSSSADFLSARNSIAOS 1014
DB 1796 SDSKSTSGST 1828

RESULT 15
B71609
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71609
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.;
Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Claydon, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Accession: B71609
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1,665 <GAP>
A:Cross-references: GB:AE001410; GB:AE001362; MID:93845245; PIDN:AACT1925.1; PID:9384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0680w

Query Match 3.6%; Score 201.5; DB 2; Length 665;
Best Local Similarity 17.8%; Pred. No. 0.0019;
Matches 135; Conservative 126; Mismatches 299; Indels 197; Gaps 28;

QY 64 YVALLOK---KDRFCSLSRIFHD-----OKKDEHASSPFVAFRMDCSKLDKL 115
DB 8 FLAVLAGFLNKKHANLINSYNDIVEDINIKKEKSSSEPPFIPIK-----IDNV 60
QY 116 KTSN-----GTAPRTLPAKQNTSGKSIYFVSTFVPAVSGOSVS 158
DB 61 HTKNNNOYNLHNKSNKTHLTGYHTSFL---QKCTINDC-----VYDKKDEIN 108
QY 159 PSTOSSGKNADRSTLPKSVQEGN-----DSKNAPSGKNGAAEANTDPMKDLG 209
DB 109 NITREKODNNNNNGT--KQIEKKKINKSDLRHONELMLOSGKNODINKNEKGDISN 166
QY 210 P-AON-YDVAAVNS-----EDNTSVDVGLPEVPOQTWHEVNGADQ 249
DB 167 SNAENKRDYKGVLELEKKKEEKISDQKVEENKSSDQKVEENKSSDKHVEENKSD 226
QY 250 PPSPTKLSVYLLKRNEDNGKTEETLVACNLTQKPNPMSKGE-----RDVYAEQN 302
DB 227 DKTIEVYKKEVEHEDEDEDEKKEKSEKKNKDEKNDENDELSDEDEVDVDEEKN 286
QY 303 ---LTKDPKPVSGQKEQICNEPCEEVYLLKSSKSKR-----TDKRLAKKQOHS 349
DB 287 ENDIDDDKRETDKTHLEEEENEILEKEFSDKKKNKDKOTKKEKSDTEKEKSDIEKE 346
QY 350 KRTAQADVSDAKLCRRKPKVRLSEIINANOVEDSRSDVYHRENAADPCEDDRSTIPV 409
DB 347 KSKDKEREKSKDKEREKDKEREKSKDI--EKEKEDKDIEREKSDTKEKEKDKDI 403
QY 410 PMEYSMDIPVSNHTVGEDGLKSSKNKTKRYSDVDGSLMNLWNGKKRTGSVHHTVA 469
DB 404 EKEKSKDM-----EKLKNKQDEKK-----NEKKKDKQDIH--- 438
QY 470 HPAGNLNKKVTPYASTOHDENDENGDLTNMHTKDVQCHVSEISTORCSSKGTAGLS 529
DB 439 -----DDNDENDMEIEENDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 478
QY 530 GKTHSAASTKYGSESTRNGONIHVLSAEQOCQMETENSVLSHAKKVPAAE---HDIOIM 586
DB 479 GNEGSENGENGENGENGENKNESENEENENENENENENENENENENENENEN 538
QY 587 SDHEOSLPKKKKKQKLEVTREKOTMIDIPMDIVELLAKQHOERQMLTETDCSDINRQ 646
DB 539 TNANKNETIKNKSEITITKS-----IDYNNRNN-----NDIDKVN 577
QY 647 SKTTADDCVIAAKGSDVYASVFDINSOOKSLASSTOKELQHLALTTOESPH---P 703
DB 578 -----NHIF--TNQOKKRNHLNHEQKFNBTLLNVSTNHNKHYEEK 614

Tue Sep 17 07:37:13 2002

us-09-828-068-2.rpr

Page 12

```

QY      704 QNFSTQ--EQOHLRMEEMVTIAAS-SPLFSHDD 735
      ::::|::|::|
Db      615 KRYESNMFNVDRMKHKNTLSMDTILHNDDKLSHKKD 651

```

Search completed: September 16, 2002, 22:25:09
Job time: 6498 sec


```

OY 267 ENGKTEETLVACQCNLTGKDPNPMGSKERDQVACQCNLTGKDPKPYSGQKCEQICNEPCERY 326
DB 115 HIGSVST -NNNSNNALIHNNPLSSH----- 139
OY 327 VLKRSKSKRKTDKLMKQOHSKSKRTAQADYDAKLCRRKPKYVLLSEIINANOVEDS 386
DB 140 -LSNPSSSLRNKSSSLVANS-----PAFADVELSKKRAVISNNMPTSMIALYQTA 191
OY 387 RDEVHREMAADPCEDR-----STTPPMEXSMOIPVSNHIVGED 427
DB 192 KSNHIGPSTASAKAFKRAFAFSNNTAPSTSNNGSNTPPAFL-----LPLPS----- 240
OY 428 GLKSKNKTKR-----RKSDVVDGSSLM--NWLNGKKKRGVSHHTVAHPAGNLSNKKY-T 481
DB 241 --LSQONKPKIIEPRMHTNRSREILGELLDDTKAKAPANST--THONGPYANDGLXI 297
OY 482 PTASTQHDENDTENGDLDTNMH-----KTQVCOHVEISITORSS-----KGKTA 526
DB 298 PHSNADNENNNKKKNNKNNKNGKNERNDTSTKCTSTKAPSTAPLIGSTDNQALTA 357
OY 527 GLSKGKTHSAATKYGESTRNGQNIHVLSAEDOCOMETENYLSHSAKVPSEHDIQIM 586
DB 358 SVSSSNADNHNKNNKKTSSNNNGNSASNTKNDIKSNMDSASTSNMNAIND----- 413
OY 587 SOLHQSILPKKKKKQ-----KLEVTREKOTMIDIDIMDIYELL-----AKNO 628
DB 414 -DSHSNSKPKTKADEFAARLATANGENISDSEETFYESAANSTKNLIFPDSSQOQO 472
OY 629 HERQJMETDSCDINRIOSKTA---DDDCVVAAKDGSVYAS-----SVFDTNSQ 676
DB 473 OQOQPPKQOQOQOQNGITSKISAPLLNNKKLLSLKNSRHISTGAILNNTATISTNPN 532
OY 677 QKSLASQSQKQLOGH-----LALTQESPHQONQSTOEOOTHLMEEMVITAASSPLF 731
DB 533 LNSNWMQNNNMLSGNHLDELSTKOEPRH-----QLOOQOPMDVQSVSDTSONP-- 585
OY 732 SHHDOYIAEAPTEHMGKDAKKLTEQOFKATTRNSPAATC-----GAQFRGQIAVDL 785
DB 586 ---DSNVIAKSPDKRSSLVLSKSPHLLSTSSNGNTISCNVATNSOLEPN---NDI 639
OY 786 TSTHWGSS-----SNVASRO-----PYIAPLDRYA-----ER 813
DB 640 STKKSLSNSTLHSSANRNSNYGDKNRPLRTVYSKIFDSNPNGAPLRRYSGVDPDHVLE 699
OY 814 AVNOYHARFEPSTIATMEASKICDRRNAGQVLYLPRESMPATHLLMMDPSLTASPNNG 873
DB 700 YIEQPH--NYPTMNSVKKDEFTNSRN-----NKFP--HJLNFYGDNNVIEEENNG 746
OY 874 TSS-----RQOMESQLNSQYAH----- 891
DB 747 DSSNNRPOHTMLOHEFIPEDNESDENDIHSFYNNHKKDLTKPLISDYGEDEVDYD 806
OY 892 -----NOYKGSTSYGSNLNGKIPL---TFEDI-----SRHQLDLHRLRPHR 934
DB 807 RPNATFNSYSGASVTHLPLHGRMPSRSNNDYDFMVGNNNTGNNQNLNE--YTPLMKKG 865
OY 935 VGVLSLQKELANNSENGC-----TQSGKLGVSIGTISHQW--NKKHFEALNSGMEFA 988
DB 866 QRHL--SRTNNSIAMSIGIHNGMDDYTHSINIINNDIVGISPHEFYSKSPFYVKV----- 918
OY 989 KWNALQLSVSS-----ADFLSARNSIAQSWTSGKGMVPLDRFQO-----DIC 1035
DB 919 --NFLYLAFVISLMTGFIIGFLLATNKELDVD-----YVMDNVATSSDELFDIT 970
OY 1036 ITNKNPADFT 1045
DB 971 VSAFNPGEFS 980

```

```

DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked nuclear protein)
DE (Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38
DE protein).
GN ATRX OR XNP OR HP1BP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98213653; PubMed=9545503;
RA Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;
RT "Comparison of the human and murine ATRX gene identifies highly
RT conserved, functionally important domains."
RL Mamm. Genome 9:400-403(1998).
RN [2]
RP SEQUENCE OF 325-1176 FROM N.A.
RX MEDLINE=97133299; PubMed=8978696;
RA le Douarin B., Nielsen A.L., Garner J.-M., Ichinose H.,
RA Jeannotin F., Loezon R., Chambon P.;
RT "A possible involvement of Tif1 alpha and Tif1 beta in the epigenetic
RT control of transcription by nuclear receptors."
RL EMBO J. 15:6701-6715(1996).
RN [3]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATR-X) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes."
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
CC - FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN.
CC - SUBUNIT: PROBABLY BINDS EZH2. BINDS ANEXIN V IN A CALCIUM AND
CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
CC SIMILARITY).
CC - SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
CC INTERACTING WITH HP1.
CC - SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC - SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB EMBL: AF026032; AAC08741.1; -
DB EMBL: X99843; CA67962.1; -
DB MGD: MGI:103067; Xnp.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; znf_rlng.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00184; RING; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW Zinc-finger.
FT ZN-FING 219 267 PHD-TYPE
FT NP_BIND 1579 1586 ATP (POTENTIAL).
FT SITE 1704 1707 DEGH BOX.

```

```

FT DOMAIN 319 322 POLY-SER.
FT DOMAIN 735 738 POLY-SER.
FT DOMAIN 1001 1004 POLY-GLU.
FT DOMAIN 1130 1135 POLY-SER.
FT DOMAIN 1182 1185 POLY-SER.
FT DOMAIN 1238 1245 POLY-ASP.
FT DOMAIN 1484 1487 POLY-GLU.
FT DOMAIN 1924 1931 POLY-SER.
FT DOMAIN 2205 2208 POLY-LYS.
FT DOMAIN 2245 2248 POLY-GLU.
FT DOMAIN 2403 2408 POLY-GLN.
SO SEQUENCE 2476 AA: 278601 MW: 90A42B790FC1FF4C CRC64:

```

```

Query Match 3.9%; Score 216.5; D1 1; Length 2476;
Best Local Similarity 19.3%; Pred. No. 0.0015;
Matches 165; Conservative 133; Mismatches 146; Indels 209; Gaps 36;

```

```

QY 83 HDQKCKDEHKASSSPFVYAKRR-WDCSKLDKLTSDNTPAPRTLPAKQNGTSDGCSIT 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 844 HERAKTQEGSSADDTGTEGROGSGCSIAAGSIEKVSIVVEFRMLCKPGVSSDGAKEP 903

QY 142 FVRSTFPAVSVGSKVSP-----STOSSGCK-----NADSTLPKSVQEGNDKCN 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 904 SVKEENV-NSEPEKRVSKTKTKTHLKRORSKRGSSGSGTDR--FPKKQSDSESSGE 960

QY 188 APSGKNAAEANTDSPKMDLOGPAQNTVAAVNEEDNTSDVGLALPEVPOITWHIEVNGA 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 961 KKQGRORGTGKKAP--DLKGETLKRQEWDSDDGE-----KLPEEEIGPFSK--GI 1012

QY 248 DQPTSTKLSVYLKRNDENGKTEETL-----VAIQCNLTQD--NPMGKERD 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1013 KQKTDTPAGGKKKKKKKDKCEKEELSDVDKLPKGJISCDSEEDKKTNRVSLREKK 1072

QY 296 QVA-----ECCNLTQDPRVSGOKCEQICNEPCE-IYVLRSSKSKRT----- 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1073 RFLSPAKSPGKRPCCSSDTEKSLKGCCDSTEKRPRIILRRNRSSKNTKEVKSAS 1132

QY 339 -----DKTLMKQOHS-KKRTAQADVDAKLCRI:PKKRYLLSEITNAQVEDSR 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1133 SSDDAEGSSSEDNKQKQRTSAKKKTGTNTEKKRNSLRAPPKRKQY-----DISSS 1184

QY 388 SD--EYIRENAADPCEDDRSTIPMEVSMIDIPVSNHT-----VGEGLKSS----- 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1185 SDIGDDQNSGESSDEOKIKP---VTEMLVLPVSHGTGQSSGDBALSKSPATVDD 1240

QY 433 -----KNKTKRKY-----SDVVDGSSSLMNLNKKKRTGSVHHTVAHPAGLS 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1241 DDDNDPKNRLAKKMLLEIKANLSSD--EDGSSDDEPDGCGKKRIQOSESPADDELIR 1298

QY 477 NKAVTPTASTQHDENDTENGLDTNMKTVCQHVSEISIQRCSSKGTAGLSKGTHTSA 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1299 REGILAVNQVNSESDSESSKKPRYRHR--LLRKLTLTSLGEGSEKPT-----KPKHKE 1352

QY 537 ASRTKGESSTRNGNIVLAEDCCQMETENSIVLSHSAKISPAEHDIQIMSDHEGSLPK 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1353 A-----KGNRRKRVSESEDDPQESGVSEE-----VSESEDEORPR 1390

QY 597 KKKKOKLEVRREKOTMIDIPMDIVELLAKNOHEROLMTITDGDINRIOS---KTADD 653
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1391 TRSAKKKLEENQRS-----YKQKKRRRIKVCYEDSSSEKKSSEEDKKEGDE 1438

QY 654 DCVIVAAKDDSDYASVFDNISOOKS-----LACQSTQKELQGHILATLTQESPH 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1439 E-----DEDEDEDEDEDENDSDSKPGKRRKIRKILKDKLLETQON--ALKEEERR 1489

QY 703 PQNFQSTQEQOOTHIRHEEMTYIAAS-----PLFSHHDDQYIAEA 742
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1490 KRLAERERERE---KLREVEIEDASPTKCPITTKILVIDENETKEPDLVQVHRMVIK 1546

QY 743 PTHMGKDKAKKLWE---QFKATTRNSPATGGAQFRPE---IQAVDLSTHYVMSSS 795
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1547 PHQ---VGVGVVMDGCCSEVSEKTKSPGSCILAHCKELGTLQVVSFLHTVLCDKL 1602

```

```

QY 796 NYASROPYIAPLD 808
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1603 DF-STALVVCPLN 1614

```

```

RESULT 4
ATRX_HUMAN
ID ATRX_HUMAN STANDARD; PRT: 2492 AA.
AC P46100; P51068; O15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chlatroni P.,
RA Colleaue L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RA Stayton C.L., Dabovic B., Gullisano M., Geetz J., Broccoli V.,
RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Geetz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKC1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RA Pearce A., Chapman J.;
RN Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP EZH2 BINDING.
RX MEDLINE=98167853; PubMed=9499421;
RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
RA Colleaue L.;
RT "Specific interaction between the XNP/ATR-X gene product and the SET
RT domain of the human EZH2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).

```

RN SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE=20040663; PubMed=10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Blackmore M.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.,
 RT "Localisation of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 RN [9]
 RP DISEASE.
 RX MEDLINE=20213147; PubMed=10751095;
 RA Villard L., Fontes M., Ades L.C., Geetz J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 RN [10]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE=97196774; PubMed=9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RT without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 RN [11]
 RP VARIANT JM GLN-2131.
 RX MEDLINE=96224392; PubMed=8630485;
 RA Villard L., Geetz J., Mattei J.-F., Fontes M., Saugier-Verber P.,
 RA Munich A., Lyonnet S.;
 RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 RN [12]
 RP VARIANTS ATR-X
 RX MEDLINE=97467722; PubMed=9326931;
 RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Azenbauer B.,
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
 RA Levin M.R., Maunio M., Neri G., Pierpont M.E., Slaney S.F.,
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 RN [13]
 RP VARIANT ATR-X LEU-246.
 RX MEDLINE=20123062; PubMed=10660327;
 RA Fichera M., Romano C., Castiglia L., Faglia P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 RN [14]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE=99347960; PubMed=10417298;
 RA Lissi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
 RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 RN [15]
 RP VARIANT CMS THR-2050.
 RX MEDLINE=99326061; PubMed=10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.;
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).
 RN [16]
 RP VARIANTS ATR-X E-175; 176-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE=99219535; PubMed=10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belongue J.,
 RA Lissi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;

RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).
 RN [17]
 RP VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
 RX MEDLINE=20451413; PubMed=10995512;
 RA Wada T., Kubota T., Fukushima Y., Satoh S.;
 RT "Molecular genetic study of Japanese patients with X-linked
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEKIN V IN A CALCIUM AND
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 CC INTERACTING WITH HP1.
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1, 2, 3, 4 (SHOWN HERE) AND 5;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
 CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
 CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
 CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
 CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
 CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
 CC SYNDROME (CMS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
 CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH
 CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME
 CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
 CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
 CC MICROGENITALISM AND EARLY DEATH.
 CC -1- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS
 CC SYNDROME (SPM). CLINICAL FEATURES INCLUDE SEVERE MENTAL
 CC RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND
 CC BILATERAL CRYPTORCHIDISM. DUE TO THE CLINICAL OVERLAP WITH ATR-X
 CC SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SPM, MIGHT
 CC BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H
 CC INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED
 CC MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE
 CC MENTAL RETARDATION WITH SPASTIC PARAPLEGIA, MICROCEPHALY, SHORT
 CC STATURE AND CRYPTORCHIDISM.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U72937; AAB49970.2; -;
 DR EMBL: U72938; AAB49971.2; -;
 DR EMBL: U72935; AAB40698.1; -;
 DR EMBL: U72904; AAB40698.1; JOINED.
 Query Match 3.88; Score 208.5; DB 1; Length 2492;
 Best Local Similarity 21.3%; Pred. No. 0.004;
 Matches 150; Conservative 102; Mismatches 286; Indels 165; Gaps 31;
 QY 62 RGVYALLQKKDKRCSLRIFHDQKDKDEHKASSPSFV-----AKFRPMQSCSLD 113
 DB 404 KAHLLAEEDLNSERPADAV-NKEKNTEKHVVDKAKETTKARKKEKPCALEKKDISSEA 462
 QY 114 KL-----KTSDNGTAPRTLPAKQNGTSDGCSIFVVR 144

```

Db 463 KLSRKQVDSSEHMHQNVTEEQRTNKSTGEHKKSDRKE:POYPEA---NTSEDLMDIYS 519
QY 145 STVPASVSGQ-----KVSPSTOSSQCKNDRSTLPKS:OEGNDSCKNAPSGKN--GAAE 197
Db 520 ---VPSPVEDIDENLETAEVSSVDHOGSSGTEQIYESSSVKNTISSKNRGIGKS 576
QY 198 ANTDSPMKDLQGPANVDYVAAVSEEDNTSVYGALPEV:OITWHIEVNGADPPSPFKLS 257
Db 577 KTLAKYTKEL-----YVKLTPLVSLRNSPTKAGDCOEYQ-----DKDGKSGGLNPKLE 625
QY 258 EYVLAKNEDENG---KTEETLVAEQCLTKDPPMPSGKI:RDQVADQCNLTKDKRPVSGQK 314
Db 626 KCGIAGENSNEHLVENEVSILTEESDLRSPVKTTPI:RRPT-----ETNPVSNIS 677
QY 315 CEQICNEPCEVYVLRKSSKSRKTKDKKIMKKQOHSKKR:AAQAVSDAKLCRRPKVY--R 372
Db 678 DEE-CHEYKE--KOKLSVPVRKKDR-----NSSDAIDNPKRPNKLPKS 719
QY 373 LSEIINANQVEDSRDE-----VHRENAADPCEDDISTIPVPEVSMIDIPVSNHTVG 425
Db 720 KQSEYVDQV-----SDSDEMLALIKGVSRMSSSSSDPT:NEIHTNKTLYDLTKQ---AG 772
QY 426 EDCIASSKSKTKRKYDYVD-----GSSIMVNLNGKRI:-----TGSVHT 467
Db 773 KD-----DKGRKRRKSTSGSDFPTKKGKSAKSIISKRI:QOTQSESSANYDELEKEIKSM 828
QY 468 VAHPAGNLNKKKVTPTPA--TQHDENDTENGIDTNNMK:DVCOHVEISITQRCSSGKTA 526
Db 829 SKTGARTTKRIRPNTKDDSSDEKHSKKGMDNCGHK-----NLKTSQCGSSDDAE 880
QY 527 GLSKGTHSAASTKYGSESTRNGQNTIHL:SAEDOCOMETENSVLASHA--KVSFAEHDIOI 585
Db 881 RKQERTFSAGETVAKDVT-----IMELRRLPKK:QAASSTGCVKLSCKE---QS 930
QY 586 MSDLHQSLPKKKKKQKLEVTREKQTMIDIPMDIYE:ILAKNQHERQMLTETDCSDINR 644
Db 931 FTSLEYRAKAEKREKSHLTKTKCKKVQDGL--SDIAEKFLKQDQD-----ETSDEDK 982
QY 645 IOSTKTADDVCYVAAKDGSDYASVFDTNSOQKSLASG:STOK 687
Db 983 KSKKGTTEE-----KKKPSDFKKVYKMEQYQS--SLGTEK 1018

RESULT 5
IGA4_HAEIN STANDARD: PRT: 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
CC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NH71 HK61;
RL MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

```

```

CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M87491; AAA24968.1;
DR MEROPS; S06.001;
DR Interpro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPRASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150ABA CRC64;

```

```

Query Match 3.6%; Score 200.5; DB 1; Length 1849;
Best Local Similarity 18.4%; Pred. No. 0.007;
Matches 176; Conservative 142; Mismatches 403; Indels 235; Gaps 34;

```

```

QY 130 KONGTSDGSIPTFVRSTFPASVGSQKSPSTOSSQCKNDRST-LPKSVQESNDSCKN 188
Db 1000 KNGQVDTNIT-----TPNDI--QADAPSAQSNNEETARVETPVPP-----A 1041
QY 189 PSCKNGAABANDNTPMKDLQGPANVDYVAAVSEEDNTSVYGALPEV:OITWHIEVNGAD 248
Db 1042 PATESAIASEQPEPTRAETAOPAMEETNTANSTETAPKSDTAQTQENP-----NSBS 1093
QY 249 OPPSTFKLESEVYLKRNEDNGTEETLVAEQCLTKDPPMPSG--ERQVAEQCLTMD 306
Db 1094 VPSST-----TEKVAENPQEN-----ETVAKNEQDAT--EPTPQNGEYAKEDQPVVEANTQTN 1145
QY 307 PKPVSGQKE-----QICNEPCEVYVLRKSSKSRKRT---DKLIMKKQOHSKKRTAAD 357
Db 1146 EATQSEKTEETQOTAEKSEPTESVTSVSENOPEKTVTSQSTEDKVVYKEKKAKVEEETQ 1205
QY 358 VSDAKLCRRKPKKRVLLSEITANQVEDSRDEVAHRENAADCEDDRSTTPVPMESMDI 417
Db 1206 KAPQVTSKEPPKQA-----EPAPREVPTDVAEBAQALQOQTPPTV-AAAE 1251
QY 418 PVSNTVGEDGLKSSKNKTKRKYSDVVDGSSLM-----NMLNGKKKRTGSVHHVVAHPA 472
Db 1252 TSPNSKPAEETQOPSEKTAIEPVTVPVSENTATQPTETETAKVEKEKQEVPOVASQES 1311
QY 473 GNLNKKVTPYASTQHDENDTENGIDTNNMK:TDVCOHVEISITQRCSSGKTAGISGK 532
Db 1312 PROEPPAARPOAQTPOAPARENVLTKNVGEPOPOAOPQOTQSTAVPTGETAANSKPA 1371
QY 533 THSAASTKYGSESTRNGQNTIHL:SAEDOCOMETENSVLASHAKVSPAEDHIDIMDILHQ 592
Db 1372 AKPOAOKAQTPAR--ENVSTVNTKEP-QSOT-----SATVSTEQPAKTTSSNV-EQ 1420
QY 593 SLPKKKKKQKLEVTREKQTMIDIPMDIYELLAKNOHERQMLTETDCSDINRISQKTAD 652
Db 1421 PAPENSINSGATITETAEKSDKO--MEVYENDROEAQTVADNSVANSSESS----- 1474
QY 653 DDCVYVAAKDGSDYASVFDTNSOQKSLASQSTQKELGHALTQOE-----SPHQ 704
Db 1475 -----ESKSRRRRSVQPKETSAEETVAASQIETTVDSVSTPKPR 1515
QY 705 NFQSQNEQOQTHLRMEEMVTIAASSPLFSHHDDQYIAEAPTEWGMKDKAKKLTWQPKATT 764
Db 1516 SRRTRSVQTN-----SYEPV-----ELPTEN----- 1537

```

```

QY 765 RNSPATCGAOPGICIAVDTSTFHVMSSSNYASROPYIADRYAERAVNOVHARNP 824
DB 1538 -----AENAVNQ-----SGNVANSQP-----ALNNTSKNTN 1566
QY 825 STIATWESKLCDDRRNAGOVVILPKESMPATHLLRMMDPSTLASEPNVGT-----SSRNOM 880
DB 1567 AVLSNMAKMAQFVALNVGAV-----SOHISOL-----EMNEGOYVWVMSNTSM 1611
QY 881 EQQLNISOYAHNKGSTSTYSGSNLNGKIPLEFEDLSHOLHDLHRLPRHPRVGLGS 940
DB 1612 -----KNKYSQYREFSKSTQTOGLMDQTS-----NNVQLGQVFTYVRNSNNEPKASS 1662
QY 941 LLOKEIANMSENCGTQSGYKLGVTGITSQNRKEH-----FEALNSG 984
DB 1663 KMTLQOVNYSKYVADNMHMYLGLDGYCKFQSNLQTNNAKFAHRTAQIGLRAGKAFNLG 1722
QY 985 MSAKMNALQQLSVSSADFLSARNSIAQSWTRGKGVHPLD---REVRODICIT 1037
DB 1723 NEAVRPTVGVRYSYLSNADFALADRI-----KVNPISTVKYAFQVLDLSTY 1768

RESULT 6
ICAL_HUMAN
ICAL_HUMAN STANDARD: PRT; 708 AA.
AC P20810;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Calpain inhibitor (Calpastatin) (Sperm BS-17 component).
GN CAST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91124109; PubMed=2577276;
RA Asada K., Ishino Y., Shimoda M., Shimoto T., Endo M., Kimizuka F.,
RA Kato I., Maki M., Hatanaka M., Murachi T.;
RT "cDNA cloning of human calpastatin: sequence homology among human,
RT pig, and rabbit calpastatins.";
RL J. Enzym. Inhib. 3:49-56(1989).
RN [2]
RP SEQUENCE OF 1-283 FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Lung;
RX MEDLINE=92235069; PubMed=1569094;
RA Lee W.-J., Ma H., Takano E., Yang H.-Q., Hatanaka M., Maki M.;
RT "Molecular diversity in amino-terminal domains of human calpastatin
RT by exon skipping.";
RL J. Biol. Chem. 267:8437-8442(1992).
RN [3]
RP SEQUENCE OF 101-317 FROM N.A.
RX MEDLINE=90037002; PubMed=2553724;
RA Maki M., Bagci H., Hamaguchi K., Ueda M., Murachi T., Hatanaka M.;
RT "Inhibition of calpain by a synthetic oligopeptide corresponding to
RT an exon of the human calpastatin gene.";
RL J. Biol. Chem. 264:18866-18869(1989).
RN [4]
RP SEQUENCE OF 125-283 FROM N.A.
RX MEDLINE=90165962; PubMed=2407243;
RA Uemori T., Shimoto T., Asada K., Asano T., Kimizuka F., Kato I.,
RA Maki M., Hatanaka M., Murachi T., Hanzawa H., Arata Y.;
RT "Characterization of a functional domain of human calpastatin.";
RL Biochem. Biophys. Res. Commun. 166:1485-1493(1990).
RN [5]
RP SEQUENCE OF 153-708 FROM N.A. (SHORT ISOFORM), AND REVISIONS.
RA Wang L.-F.;
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 523-708 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95038502; PubMed=7951045;
RA Wang L.-F., Wei S.-G., Miao S.-Y., Liu Q.-Y., Koide S.-S.;

```

```

RT "Calpastatin gene in human testis.";
RL Biochem. Mol. Biol. Int. 33:245-252(1994).
RN [7]
RP SEQUENCE OF 425-708 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95221646; PubMed=7706496;
RA Despres N., Talbot G., Plouffe B., Boire G., Menard H.A.;
RT "Detection and expression of a cDNA clone that encodes a polypeptide
RT containing two inhibitory domains of human calpastatin and its
RT recognition by rheumatoid arthritis sera.";
RL J. Clin. Invest. 95:1891-1896(1995).
RN [8]
RP SEQUENCE OF 244-708 FROM N.A.
RC TISSUE=Placenta;
RA El-Amine N., Talbot G., Despres N., Asselin C., Boire G.,
RA Menard H.A.;
RT Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [9]
RP PHOSPHORYLATION.
RX MEDLINE=91139659; PubMed=1995645;
RA Adachi Y., Ishida-Takahashi A., Takahashi C., Takano E., Murachi T.,
RA Hatanaka M.;
RT "Phosphorylation and subcellular distribution of calpastatin in human
RT hematopoietic system cells.";
RL J. Biol. Chem. 266:3968-3972(1991).
CC -1- FUNCTION: SPECIFIC INHIBITION OF CALPAIN (CALCIUM-DEPENDENT
CC OF MEAT AND HAVE BEEN HYPOTHESIZED TO BE INVOLVED IN MUSCLE
CC CYSTEINE PROTEASE). PLAYS A KEY ROLE IN POSTMORTEM TENDERIZATION
CC PROTEIN DEGRADATION IN LIVING TISSUE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: HAS FOUR INHIBITORY DOMAINS.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- PTM: PHOSPHORYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D16217; BAA03747.1; -
DR EMBL: D50827; BAA09438.1; -
DR EMBL: M86258; AAB59398.1; -
DR EMBL: M82230; AAA52066.1; -
DR EMBL: M28227; AAA52066.1; JOINED.
DR EMBL: M28228; AAA52066.1; JOINED.
DR EMBL: M28229; AAA52066.1; JOINED.
DR EMBL: M33328; AAA52296.1; -
DR EMBL: Q26724; AAC50136.2; -
DR EMBL: S73329; AAB32311.1; -
DR EMBL: U38525; AAA80684.1; -
DR EMBL: U31345; AAB60371.1; -
DR EMBL: U31346; AAB60372.1; -
DR PIR: A34428; A34428.
DR MIM: 114090; -
DR InterPro: IPR001259; Calpain_inh1b.
DR Pfam: PF00748; Calpain_inh1b; 4.
KW Repeat; Thiol protease inhibitor; Alternative splicing;
KW Phosphorylation.
FT DOMAIN 1 169
FT REPEAT 170 222
FT REPEAT 304 356
FT REPEAT 446 499
FT REPEAT 583 636
FT VARSPLIC 212 224
FT VARIANT 592 592
FT CONFLICT 467 467
FT CONFLICT 486 486
FT CONFLICT 543 543
L.
INHIBITORY DOMAIN 1.
INHIBITORY DOMAIN 2.
INHIBITORY DOMAIN 3.
INHIBITORY DOMAIN 4.
MISSING (IN SHORT ISOFORM).
G -> E.
/FTID=VAR_005298.
R -> L (IN REF. 5).
VRD -> GRE (IN REF. 5).
V -> L (IN REF. 5).

```

FT CONFLICT 562 562 MISSING (IN REF. 8).
 SQ SEQUENCE 708 AA; 76484 MW; CACD755C928413EA CRC64;

Query Match

Best Local Similarity 3.6%; Score 197; DB 1; Length 708;
 Matches 155; Conservative 109; Mismatches 284; Indels 216; Gaps 35;

QY 70 KKPFCSSS---RIPHDK---KCDH-----KASSPFSVAKFRMDCSKCLDL 115
 DB 32 KPEPEKSSSTKLVSVEHKEGSGEGRKTEPEKSPKQFSDGSDNAHKKAVSRSAEQ- 90
 QY 116 KTSNDCTAPRTLPKONGTSDGCSITFVSTVPASVGEQKSPSTQSSOGKNAADSTLP 175
 DB 91 OSEKSEKTEKTRKPRDM-----ISAGEESVAGITAIIS-GRPGDKKKK 131
 QY 176 KSVQES--NDSCNAPSGKNGAANEANTDSPMDLOGPACNYVAAVNSDNTSYDVALP 233
 DB 132 KSLTPRAVPEKRPDKPSGSG-MDAALDLDLIDLGPEE-----TEBENTTY--TGP 180
 QY 234 EV--PQITWHIEVNGADOPPSTPKLSEVYLKR----- 263
 DB 181 EYSDPMSTYIEELGKRETIPTPKYRELLAKKEGITTGPADSSKPIGPDIDAISDFT 240
 QY 264 -----MEDNGKTEETLVAEQCNLTQDPMMSGKER-----DQVABQCNLT--T 304
 DB 241 CGSPFAAGKTEKEESTEVLLKQASAGYVSAAPPOEKKRKVKEDTMSQDALSLASIGT 300
 QY 305 KDPKP-----VSGQCEQICNEPCEVYLK-RSSKSKRTDKMKMKQOHS 349
 DB 301 ROAEPELDLSIKYEDEAKAKKEKLEK-CGEDETIPESEYRLKPTVDGKFLPEPEK 359
 QY 350 KKRQAGVY-----SPAKLCKRPKPVYRLISETITANVYEDSR-----SDEVHREN 395
 DB 360 PPRSESELIDELSEDFRSECKEPRK-----PTEKTEESKAAPAPVSEAVSRIS 411
 QY 396 AADPCEDRSTIPVMEYSMDIPVNHVTEGDEGLSKSKNTKRRKSDVDVDDSSLMNLTN 455
 DB 412 MC-----SIOGAPREPATILKGTVP-----DDAYEALADSLGKKEAD-PEDEGKPVMDVK 459
 QY 456 GKKK-----RFGSVHNYV-----AHPAGNLNKKVTPPA3TOH-----DDE 491
 DB 460 EKAKEDREKLGKEKETIPDYRLLEVVDKDKGKPLLPKE3KEQLPRMSEDFLLALSDIF 519
 QY 492 NDTEGLDITNMKTDVCOHVEISITQRCSSKGTAGLSK3THSAASKYKGESTRNCQN 551
 DB 520 SGPOWASSLKEDAKLAALISEVVSQ-----PPASTTOAGAPPRD--- 559
 QY 552 IHVLSAEDQOMETENSVLSHS-AKVSPAEDHDIQMSDLHEQSLPRKKKKOKLEVTREKQ 610
 DB 560 -----TSQSDKDLDAIDKLSLQRODPDENKPMGD-----KYKEKAKAE-HRDKL 607
 QY 611 TMIDD-IPMDIVELLAKNQHROJMTETDCSDINRIOSK7TADDVCYIVAAKDSDYASS 669
 DB 608 GERDRTIPPEYRHLIDNGDGPVKPPTKKE-----DSKIPADQDPIDALSGLDSCPS 663
 QY 670 VFDINSQ-----KSLASOSTQKELQGLH---ALTQESVHPQN 705
 DB 664 TTETSOHTAKDKCKKAASSSKAPNGKAKADSAKTETETIKPKD 707

RESULT 7

MAPB_HUMAN ID MAPB_HUMAN STANDARD; PRT; 2468 AA.
 AC P46821;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1B (Map 1b) [contains: MAP1 light chain LC1].
 GN MAP1B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=95104835; PubMed=7806212;
 RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
 RT "Cloning of human microtubule-associated protein 1B and the
 RT identification of a related gene on chromosome 15.";
 RL Genomics 22:273-280(1994).
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
 CC from MAP1B by proteolytic processing. It is free to associate with
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region
 CC of MAP1B (by similarity).
 CC -1- SIMILARITY: TO MAP1A.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L06237; AAA18904.1; -
 DR MIM: 157129; -
 DR InterPro: IPR00102; MAP1B_neuraxin.
 DR Pfam: PF00414; MAP1B_neuraxin; 10.
 DR PROSITE: PS00230; MAP1B_NEURAXIN; 6.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
 FT REPEAT 1878 1894 MAP1B 1.
 FT REPEAT 1895 1911 MAP1B 2.
 FT REPEAT 1912 1928 MAP1B 3.
 FT REPEAT 1929 1945 MAP1B 4.
 FT REPEAT 1946 1962 MAP1B 5.
 FT REPEAT 1963 1979 MAP1B 6.
 FT REPEAT 1997 2013 MAP1B 7.
 FT REPEAT 2014 2030 MAP1B 8.
 FT REPEAT 2031 2047 MAP1B 9.
 FT REPEAT 2048 2064 MAP1B 10.
 FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 KKEE AND KKEI/V REPEATS).

Query Match 3.6%; Score 196.5; DB 1; Length 2468;
 Best Local Similarity 18.4%; Pred. No. 0.016;
 Matches 204; Conservative 170; Mismatches 448; Indels 289; Gaps 44;

QY 167 KNADRTLPKSVQEGNDSCNAPSGKNGAANEANTD--PKKDLOGPAQNVVAAVNSDNT 224
 DB 506 KHHDFLKQPLATQKDLTGVPYVVKOTLKQADRESLSKPAKFLPKSVKESKE-- 563
 QY 225 TSDVAGALPEVQITWHIEVNGADOPPSTPKLSEVYLKRNEDNGKTEETLVAEQCNLT 284
 DB 564 -----ETPEVT--KVNHEKPKVSKSEKAVWKDKPVKTETKPSVTEKEVPSKE 611
 QY 285 DPNPMGKERDQVAEQCNLTQDKPVPVSG-----QKCEQICNEPCEVYLKRSSKSKRTKD 339
 DB 612 EEPSPV-----KAEVAER--QATDVKPKAAKEKTYKKTETKVPEDKKEKPKKEVAKKED 665


```

DR PROSITE; PS50088; ANK_REPEAT; 20.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation.
FT REPEAT 63 92 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 129 158 ANK 3.
FT REPEAT 162 191 ANK 4.
FT REPEAT 193 220 ANK 5.
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A.
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A.
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH.
FT VARSPLIC 1039 1039 O -> OFGLKHLPLAPPLNMGESLVSRIILQLGPPGTX
      (IN ISOFORM 2).
FT VARSPLIC 1444 3528 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT CONFLICT 475 476 GO -> PE (IN REF. 4).
FT CONFLICT 971 971 I -> S (IN REF. 1).
FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).
FT CONFLICT 3586 3586 I -> Y (IN REF. 1).
SO SEQUENCE 3924 AA; 430337 MW; 52AC496C428329D2 CRC64;

```

Query Match 3.5%; Score 194.5; D3 1; Length 3924;
 Best Local Similarity 18.7%; Pred. No. 0.038;
 Matches 199; Conservative 167; Mismatches 391; Indels 309; Gaps 46;

```

QY 28 VAPVIELTATPRQDAAGVDEPAQH-----QCEHPSIRGVALLQKKPKFCSL 79
DB 2517 VTPKTTDVSTKPAVHICAEDESENCKRFTPEEEMPKWTKIKMFDELEQ-ANK 2575
QY 80 RIFHDQKCDHKKASSPFSVAKFRMDCSKCLDKLKT(S)NGTAPRTLPANONGISDCS 139
DB 2576 RDIYKKEPQESSSSSDPA-----DCSVADVEPKHTISGEDESGVP----- 2617
QY 140 ITFVRSITFPAVSQKVSPTOSQ-----GKNADRSTLPKSVQEGNSKCNAPSGKNG 194
DB 2618 -----VLVTSESRKVSSESEPELQJLKGADSLPPEV----- 2653
QY 195 AAEANTSPKMDLOGPRAONTVAANVEDNTSVGVGLPIVQITWHTIEVNCADQPPSTP 254

```

```

DB 2654 -IRVPPSLP-----SSMDSNSPEENQF-----QPVG- 2682
QY 255 KISEVVLAKRNE-----ENGTEETLVAEQNLTKDPNMSGKERDQVAEQCNL-TKDPKPV 310
DB 2683 --KQYTFKKNEDTQEEPRGKSEEEKDSES--HLAEDRAHASTEADRSYDKLNDTDQPKIC 2739
QY 311 SGQCKQIC-----NEPCEEVVLKRSKSKRKDKLMMKQO--HSK 351
DB 2740 DGHGCEAMPSSSARPVSSGLOSPTGDVDEQPVYIKESLALQGHKEKTEEEELDVSA 2799
QY 352 RTAQADV-----SDAKLCRRKPKVRLSEIINA-----NOVEDSRDEVHENAADPE 401
DB 2800 ESPQACCPSEPSFSSSSSLPHCLVSEKELDEDISATSSQKTEVYTKTDTF--ENLPKDP 2858
QY 402 DDRSTIPVME--VSMIDPVSN-----HTVGEDGIKSKNKT----- 437
DB 2859 SODSSITQTDRFSDVDPVSDLAENDEIVDQITSPYENVPSQSFSSSESKTQTDANTH 2918
QY 438 -----RKYS-----DYVDGSSLMNMLNGK-----KRGSVHHTVAHPAGNIS 476
DB 2919 TGFHSSEVYSVITTSPEVDVVAASSSGTVLSKESNFEGQDIKMSQLESTLWEMQSDSV 2978
QY 477 NKKVPTAS-----TOHDENDT-----ENGLDTNMKHTDVCQVSEI 514
DB 2979 SSSFEPTMSATTTVYGEQISKVILITKTVDSDSMSREDDDAFERVKEE--QKIFGL 3036
QY 515 STQRCSSKKTAGLKGKTHSAASTKYGGESTRNGONIHVLSAEDQCOMETENSV-LSHS 573
DB 3037 MVDK-OSQGTPTDTPAPRPTTEGT-----PTSEGNPLFOEGKLFETRSGAIDMTFR 3089
QY 574 AKVSPAENHIQIMSDLHEOSLPKKKKKOKLEYTREKQMTIDIDIVYELLAKNGHERQL 633
DB 3090 STADESFHFQIGQESRETL-----EDYKKGATGADPLP----- 3126
QY 634 MTEDCSDINRIQSKTTADDDCVIY--AAKDGSDYASVPFTNSQKSLASQSTOKELQ 690
DB 3127 --ETSAESIALSESKETVDEADLLPDVSSEVEEIPASDAQLNSQMG--IASSTETPK 3182
QY 691 GHLALTQESHPQ-----NFQSTQEOOTHKREMYNTIANS-----SP 729
DB 3183 EAVSVGTDLPTVQGDIPPLSGVKQISCPDSEPAVQVQ-LDFSTLRVSYSRGGDSP 3241
QY 730 LFSHDDQYIAEAPTEHMGKRDKAKLTWEQFATRRNSPAT--CGAOFRGIOAVDIT 786
DB 3242 DSSPEEQKSVIEIIF--APMENVPFTESKSIIPRTMPTSTRPAPSAEYESVSEDEL 3298
QY 787 STHVNGSSSNTA-----SRQVIAPLDRYERAVNOYHARNPSTIATM--EASKLCD 837
DB 3299 SV-----DEENKADKAPKSKLPEVKVPLQR-VQOQLSDIDT-SVQKTVAPQGDMAIAPD 3352
QY 838 RNACQVVLVPRESNPATHLRLMMPSTLASPPNPGTSSRNOMESQLNNSQYAH--NQ 893
DB 3353 NKSKE-----SDASSLDSKTKCPVTRSTVETETESRERAELELES 3395
QY 894 YKGSF-----STVGSNUNGKIPLTFEDLSRHQHLHR 927
DB 3396 EGGATRPILNLSRLPVKSRSTTSSCRGCTSPKE--SKHEFFDLR 3439

```

```

RESULT 9
YF16_YEAST
ID YF16_YEAST STANDARD; PRT: 1233 AA.
AC P43597;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 137.7 kDa protein in UGS1-FAB1 intergenic region.
GN YF016C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]

```


RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA MEDLINE-95400292; PubMed-7670463;
 RA Murkani Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasunuma S.-I., Sasunuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT Analysis of the nucleotide sequence of chromosome VI from
 RT *Saccharomyces cerevisiae*;
 RL Nat. Genet. 10:261-268(1995).
 CC -1- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL: D50617; BAA09255.1; -
 DR SGD: S0001912; YPR016C.
 KW Hypothetical protein.
 SQ SEQUENCE 1233 AA; 137697 MW; C8AYCD26F0892P6 CRC64;

Query Match 3.5%; Score 191; DB 1; Length 1233;

Best Local Similarity 18.9%; Pred. No. 0.013; Matches 385; Indels 318; Gaps 48;

Db 360 EKGQTESDSDGDISPSVLANKQKTEIGKEDHVEQKDE----- 400
 QY 50 EPAHQCEHFSIRGVALLQKDPKCSLRIPHDKCKDEHKASSPFSVAKFRMDCS 109
 QY 110 KCLDKTKT-----SDNGTAPRTL-----FAK-----QNGTSD 136
 Db 401 KCKEKLTVHNNMHNMFNAAGSDSIIPETETETVDEDTMGFTKRIIDNEKMLQHGTDN 460
 QY 137 GCSITFVR-----STFVPASVGSQKVSPPSSQSGKNADSTLPSQVQEGDSK 185
 QY 461 -ISVEVEKEEEEEEEENSTF--SKYKENVY--GEDEAVRNNNEVSTEEESTSKGEE-- 514
 QY 186 CNAPSGKNGAENATDSPKMDLOGPANTDVANVEDNTSVVGLAPEVP--QITWHIEV 244
 Db 515 --IMGDEKQSEAGEKSSITIEGSA--SAKISKDNVLIDEA--EAPQENKRTVEV 566
 QY 245 NGADQPPSPKLSVYVLRKNEDENGKTEETVLVAEQCNLTDPHPSGKEKEDQVAEQCNLT 304
 Db 567 VGEIDIPADPRDVEIVE-----AVEKNITPEDLEVAKE-----DOEGEYKLD 610
 QY 305 KDPKPYSGQCEQICNEPCEEVVLRKSSKSRKTDKLMKKQSHSKRTQAQVSDAKLC 364
 Db 611 EPYKAKMKDKIAMRGAESEISDEMKKQEGTAEISNEKAKKEVETARESDVGEVEKST 670
 QY 365 RRRPKRVRLSEITIANQVED---SRSDVHRENAADPCEDDRSTIPVPEVSM---- 416
 Db 671 PESPKYVVKRCT---SGREDQLINERDPEVLKEDVAVPDEDVKKPEIATITENSEEDPK 726
 QY 417 ---IPVSNH-----TVGEDGLKSS--KNTKRRYSVVVDGSS--MNMVNGKKKRT 461
 Db 727 SQRVQVSTQEAETQKDMGVGVSTTSFKEKEKPRREIYQEGKTIQKDTNHEGTEEA 766
 QY 462 GSVHHVHAPAGNLSNKKVTPA--STOHDENDTENGIDTNMKTVDVCOHVEISTQRCS 520
 Db 787 AS-ENSKASDVGT--ARKYIEPSSSESVKKTDEEAE--VENSEKTEIRIKYKAELENIDAP 841
 QY 521 SKCK--TAGLSKGTTHSAASTKYGESTRNGQNIHVLSAEDQCMETENSVALSHSAKVS 579
 Db 842 KEAEVYAELENK-----ENEDVEVDTEED---AEVENSEKTEIFIKYK-- 879
 QY 580 EHDIQMSDLHEQSLPKKKKKKQKLEVTREKQTMIDIPMIVELKAKHQRQLMPTDC 639
 Db 880 -----AEIGNIDAPK-----EAETVYAELENKEDN-----VEVAATSKED---ITKC 918

QY 640 SDINRIQSKTTADD-----DCVIYAKD-----GSDYASVPEDTNSQO----- 677
 Db 919 SE-----PAETPIEDGCTEAEVSKKDAEAVTKEDENMENSIAEALDVGDDEIDINI 974
 QY 678 -----KSLASQTOKELO--CHLALITQESHPNPFQSTOE 711
 Db 975 SDEFQRTVELPELEKODIKNKGEDKLEVEFEKETSLPDLVVEENITEKNEIKQEE 1034
 QY 712 QOHLRKEENVTIAASSPLFHHDDOYIAEAPPEHMKRAKALT----- 756
 Db 1035 EVSOLDPNEFESISKEAP--NNDENGFEDQST---RENPKRASADIDFKILDETNEF 1087
 QY 757 WEQFKATTRNSPAATGCAQFRPGIAVDLSTVHMGSSSVYASQPIAPLDYAEAAVN 816
 Db 1088 LEQKTIYDSELAN-----LQSLDAKDSITQTEGSKKNNDKPDY----- 1129
 QY 817 QVHARNEPSTIATMEASKLCDRNAGOVVLYPESMPATHLRMDPSTLASPPNYGTSS 876
 Db 1130 -----ITTSEIRKLEKE---PVYIYTSIAGGCFHMI---PRT----- 1161
 QY 877 RNQMESQLHNSQYAHNOYKSTSTVGSNLANGKTIPLTFEOLSRQLHDLRPLRPHRVG 936
 Db 1162 -NRL-----STILYAN--RIPFYRLDGTDD--EARKVMTESKGR 1197
 QY 937 VLGSLQKE---IANNSENGCTQSGYKL 961
 Db 1198 SLPGVARGHNDLIGNWEIEEANDYKL 1225

RESULT 10

ID RPL_HUMAN STANDARD; PRT; 2156 AA.

AC P56715, 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Oxygen-regulated protein 1 (Retinitis pigmentosa Rpl protein)

DE (Retinitis pigmentosa 1 protein).

GN RPL OR ORP1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

ON NCBI_TaxID=9606;

RN 11 SEQUENCE FROM N.A., AND VARIANTS HIS-872; TYR-985; THR-1670; PRO-1691

RP AND TYR-2033.

RC TISSUE-Retina;

RX MEDLINE-99318095; PubMed-10391212;

RT Pierce E.A., Quinn T., Meehan T., McGee T.L., Berson E.L., Dryja T.P.;

RT "Mutations in a gene encoding a new oxygen-regulated photoreceptor

RT protein cause dominant retinitis pigmentosa.";

RL Nat. Genet. 22:248-254(1999).

RN 13

RP SEQUENCE FROM N.A.

RX MEDLINE-99310563; PubMed-10401003;

RA Guillonneau X., Parlev N.I., Danciger M., Kozak C.A., Cideciyan A.V.,

RA Jacobson S.G., Farber D.B.;

RT "A nonsense mutation in a novel gene is associated with retinitis

RT pigmentosa in a family linked to the Rpl locus.";

RL Hum. Mol. Genet. 8:1541-1546(1999).

CC -1- FUNCTION: COULD HAVE A ROLE IN THE DIFFERENTIATION OF

CC PHOTORECEPTOR CELLS.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN RETINA. NOT EXPRESSED IN HEART,

CC BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, SPLEEN AND

CC PANCREAS.
 CC -1- DISEASE: DEFECTS IN RPI CAUSE RETINITIS PIGMENTOSA FORM 1 (RPI). A
 CC DISEASE CHARACTERIZED BY CONSTRUCTION OF THE VISUAL FIELDS, NIGHT
 CC BLINDNESS AND FUNDUS CHANGES. THE DISEASE SEEMS TO BE ASSOCIATED
 CC WITH TRUNCATED (STOP OR FRAMESHIFT MUTATIONS) FORMS OF THE
 CC PROTEIN.
 CC -1- SIMILARITY: CONTAINS 2 DOUBLECORTIN DOMAINS.
 CC -1- DATABASE: NAME-Retnet;
 CC NAME-Retinal information network;
 CC WWW="http://www.sph.uth.tmc.edu/retnet/";
 CC -1- DATABASE: NAME-Mutations of the RPI gene;
 CC NOTE-Retina International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/rpimut.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb.ch/announce/
 CC or send an email to license@isb.sib.ch).
 CC -----
 DR EMBL: AF143226: AADA4197.1: -
 DR EMBL: AF143224: AADA4197.1: JOINED.
 DR EMBL: AF143225: AADA4197.1: JOINED.
 DR EMBL: AF143222: AADA4198.1: -
 DR EMBL: AF141021: AADA2072.1: -
 DR EMBL: AF152242: AADA6774.1: -
 DR EMBL: AF152240: AADA6774.1: JOINED.
 DR EMBL: AF152241: AADA6774.1: JOINED.
 DR EMBL: AF146592: AADA6769.1: -
 DR MIM: 603937: -
 DR MIM: 180100: -
 DR InterPro: IPR003533: DCX.
 DR SMART: SM00537: DCX; 2.
 DR PROSITE: PS0309: DC; 2.
 KW Vision: Retinitis pigmentosa; Retinal protein; Polymorphism; Repeat.
 FT DOMAIN 36 118 DOUBLECORTIN 1.
 FT DOMAIN 154 233 DOUBLECORTIN 2.
 FT DOMAIN 268 273 POLY-SER.
 FT DOMAIN 671 675 POLY-LYS.
 FT DOMAIN 1687 1691 POLY-SER.
 FT VARIANT 872 872 R -> H.
 FT VARIANT 985 985 /FTID=VAR_007810.
 FT VARIANT 1670 1670 N -> Y.
 FT VARIANT 1691 1691 /FTID=VAR_007811.
 FT VARIANT 1691 1691 A -> T.
 FT VARIANT 2033 2033 /FTID=VAR_007812.
 FT VARIANT 2033 2033 S -> P.
 FT VARIANT 2033 2033 /FTID=VAR_007813.
 FT VARIANT 2033 2033 C -> Y.
 FT VARIANT 2033 2033 /FTID=VAR_007814.
 SQ SEQUENCE 2156 AA; 240659 MW; 55A8DECA3DFA507 CRC64;

Query Match 3.4%; Score 190.5; DE 1; Length 2156;
 Best Local Similarity 20.3%; Pred. No. 0.028; Indels 291; Gaps 52;
 Matches 204; Conservative 136; Mismatches 373;

QY 93 ASSPESVAKFRMDCKLDIKTSNGTAPRTLPKAKOQGTSDGCSITFVRSFVSPASY 152
 DB 222 AGREFP---KPGNDIDIKYL--LPARLPGISORYP-KGAKSSKRI-----SIHMSSSS 271
 QY 153 GSQKVSSTQSSQCKNA--DRSTLPKS--VOEGNDSCKNPPSGKNGAANEANTDSPMKDQ 208
 DB 272 RSQIYSVSSSEKTHNDDCYLDYSFEPKYLALFNDSQ-NLPI-----YPSDD-- 317
 QY 209 GPAQNDVVAANV--SEDTNVDVAGALPEVQITWHIEVNAADQPSPTPKLSEVYLK---R 263
 DB 318 -----DIKSIIFNDQGT-----MVEKMKVRRIRKEEETIKWTIVSKTGPS 359
 QY 264 NEDEBN-----GTEET--LVAEQCNLTQDPPM--SGKE-----RDQVAEQ 300

DB 360 NNDKSEMSFPGRTESRSSGLKLAACSFADVSPMERSSNQGSIAEINIQTQVAET 419
 QY 301 C-----NLTKPKRPVSG---OKCEQICNEPCEVYLKSSK-----SKRTDCKL 342
 DB 420 CSSASSEMNTVDTDIIQGDQDAKHFFYPFPGLRRNOKKSVGSLVLSVETVQEM 479
 QY 343 MKKQOHSKRTQAADVSDAKICRRPKRYR-----LSEITANQVEDSRSDVHRENA 396
 DB 480 ICQFSYSEEREGSEKNEVHMFTSCSKMSVSNKPVLYQINNNDQMESSIERKESL 539
 QY 397 ADPCEDRSTIYVPEVMSNDIVSNHTVGEOLKSKSKTKRKYSDV-----DDGSSLM 451
 DB 540 LKSSAISACVITTSOKMKEMSHN-----GLPSTISNLSIVEDVYDVCVLDKTKG 593
 QY 452 MWLNKKRKRTHVHTVAPAGNLSKKVTP--TASTQHDNDENGDLTMMHKTKDVCQH 510
 DB 594 NF-----KTYGN-----TNDPSPISADATFFSSNN--SGTDKNISERPASEA 634
 QY 511 VSEISTQ--RCSSKRTAGLSK-----GKTHSASTKYGGESTRNGONIHVLSAEDQCM 563
 DB 635 SSTVARDRLINEPAQGLTKLPKNEKKILSVASAKKKKSRQDAIN-----SRYODGL 690
 QY 564 ETENSVLSSAKV-----SPAENDIQIMSDLHQSLEPKKKKKKLEVTYR 607
 DB 691 ATK-GILKNKERINKGRTKEMLYQDSDFPKGGLCEEDL-----OKSDTVI 738
 QY 608 EKQIMIDDPMDIYELLAKNOHERQ-----LMTETDCSDINRI-----QSKTTA 651
 DB 739 ESNFPCS--KSNLNSSTIKNFRNKLNTQNSKVOGLLTFRKRSKLKISLGAPKKKEIG 796
 QY 652 DDDCVYAKGSDYASVFEPTNS-----QOKSLASQSTQKELGHALTQESBHP 703
 DB 797 QRDYVF--PHNESYCKSTFENKSLFHFENILEOK-----P 830
 QY 704 QNPSTQEQOQTHLMEEMVYTAASSPLFSHHDDQYIAEAPTEHMGKRDQAKLTQEQKAT 763
 DB 831 KDFYAPQQAQ-----VASGYLRGAKKSLVSKYVDSHTLLKSOKKKRQKQKAKAS 880
 QY 764 T-RNSPATCGAQRPGIOAVDLTSTHVGMSSSNVYASROPYIADRYAARAVNOVHARN 822
 DB 881 AILSKQAHATTRANSIASIKRDPPEALAHNSIQNYI--QSLQININPY----- 926
 QY 823 FPTIATMEASKLCDRRNAGQVLYLPKESMPATHLLMMNPSTIASPPNYGTSRRNOMES 882
 DB 927 --PTLKRIKAPVC--KNETSIVCNSNFSGN-----DPHT-----NSGKISNFVMS 971
 QY 883 QLNSQYAHNOYKGSTSTYGSNL--NGKILPTEEDLSRHQLDLRPLRPHRVGLGS 940
 DB 972 NKHITKIA-----GLT-----GDNLCKEGDKSFANDTGEEDLHETQ-----VGS 1011
 QY 941 LQKEIANWSENGCTOSGYKLGVSTGITSQNMKKEFEALNSG 984
 DB 1012 LNDAYIVPLHEHC-TLS-----QSAINDH--NTKSHIAEKSG 1046

RESULT 11
 MAPB_RAT STANDARD: PRT; 2459 AA.
 ID MAPB_RAT P15205: 062958; 09821; 09092;
 AC 01-APR-1990 (Rel. 14, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1b) (Neuraxin) [Contains: MAP1
 DE light chain LC1].
 GN MAP1B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN NCBI_Taxid=10116;
 RX [1]
 RP SEQUENCE OF 1-142 FROM N.A.
 RC STRAIN=SPRAGUE-DANLEY; TISSUE=Nestis;
 RK MEDLINE=96257242; PubMed=8666293;


```

OY 749 RKDAKLTWEQFATTRNSPATCGA-----OFRP-----GIQAV 783
Db 1178 LSGSKT--DATGDRDYNASASTISPPSSMEEDKFSK3ALRDQVREEDYKTAELDIX 1235
OY 784 DLSTVHWGSSSVYASQPIATLDR--YAEKRVNVOYARNPSTIATMEASKLDCRRNA 841
Db 1236 DVSDEKLSPAKSSLSLSPSP--SPIETKPLGERSVNFSLTPNEIKASAEGETAVVS--PGV 1293
OY 842 GQVVLVPKESMPATHLIRMDPST---LASFPNYGT:SRNOMEQL-----HNSQ---- 888
Db 1294 TQAVVEHCASPEKTELEVYSGSVTGSAGHTPYQ:NPTEKSSHLPTVEENAOAPV 1353
OY 889 -YAHNQKSTSTSYGSNNGKIPLEFDLSRHQJLHDJRRPLRPHRYVGLG-----SL 941
Db 1354 SEFTTEAKENERRSISPMDEPVP-----DSESPTEKYLSPRLRSPPLIGSEAYEDFLSA 1408
OY 942 LQKELIAMSENCQSGYVLGYSTGITSQNMKKEFFIALNSGMSAKNALQLGVS--- 998
Db 1409 DOKALGRSESPBEKNGKQFS-----DKESPVC:DLTSIDLXQDKQEKRAQFIPIK 1460
OY 999 -----SSADPLSARNSIA 1012
Db 1461 EDFSPKKASDAEIMSSOSALA 1482

RESULT 12
ZIPL_YEAST STANDARD; PRT; 875 AA.
ID ZIPL_YEAST
AC P31111;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Synaptonemal complex protein ZIPL.
GN ZIPL OR YDR285W OR D9819.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BR1824-3B;
RX MEDLINE=93161412; PubMed=7916652;
RA Sym M., Engelbrecht J.A., Roeder G.S.;
RT "Zipl is a synaptonemal complex protein required for meiotic
RT chromosome synapsis."
RL Cell 72:365-378(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Tatch A., Trevisan E., Vignati D.,
RA Wilcox L., Woldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (Mar-1996) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: REQUIRED FOR MEIOTIC CHROMOSOME SYNAPSIS AND CELL CYCLE
CC PROGRESSION. MAY ACT AS A MOLECULAR ZIPPER TO BRING HOMOLOGOUS
CC CHROMOSOMES IN CLOSE APPPOSITION. ZIPL MAY ENCODE THE TRANSVERSE
CC FILAMENTS OF THE SYNAPTONEMAL COMPLEX.
CC -I- SUBCELLULAR LOCATION: SYNAPSED MEIOTIC CHROMOSOMES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; L06487; AAA35239.1;

```

```

DR EMBL: U51031; AA864474.1; -.
DR PIR: A45173; A45173.
DR PIR: S30868; S30868.
DR SGD: S0002693; ZIPL.
KW Nuclear protein; Meiosis; Coiled coil.
FT DOMAIN 177 333 COILED COIL (POTENTIAL).
FT DOMAIN 397 438 COILED COIL (POTENTIAL).
FT DOMAIN 456 752 COILED COIL (POTENTIAL).
FT COMPLECT 55 55 T -> A (IN REF. 1).
SQ SEQUENCE 875 AA; 100035 MW; 674F12625CC9D9DFD CRC64;

Query Match 3.3%, Score 183; DB 1; Length 875;
Best Local Similarity 19.3%; Pred. No. 0.022;
Matches 161; Conservative 149; Mismatches 362; Indels 162; Gaps 34;

OY 129 AKONTSQSGCSTTFPRSTFFVAPASGQKVPSTQSSQCKNADRSTLPSVQEGDSKCN 188
Db 109 AIENDTDDEFTTEVREV-----SEGVAKETKESHDPNDSETTLK-----DKMHE 155
OY 189 PSKNGAA-----EANTDSPMKDLOGPAQNYDVANVSEDNTSYDVALPEYQJLTH 241
Db 156 YTMNGKAPLHTSINNSSTSSNDVLEAFNTQRTGCSNLQK-----LQKQONAK 207
OY 242 IEVNGADQPPSTPKLSEVYLKRNEDENGKTE---ETLVAECNLTQDPNPSGKERDQVA 298
Db 208 LKVALQSYASNSDKINEV-----GKYKSCLETLOERLATLTSHKNNQETKLKDL-- 257
OY 299 EOCNLTQDPKPVSGQCKE-QICNEPCEEVLKRSKSKRTDKKIMKQOOSKRTQAD 357
Db 258 -RQNHQLYQRISGFKSIEMLKNTINDL-----GKKKKADAFIMKKGIEYELKRELD 311
OY 358 VSDKCLKRRPKKRYLLSEILMANQVDSRS-----DEVHEENADPCEDDRSTIPVP 410
Db 312 DCSGLSEEEKIKNSSLQK-MGNKREMIKSIENFSEDKAHH----- 353
OY 411 MEVSMIDIPVSNHTVGEDGLKSSKNKTRKYSQDVVDGSSLMNMLNGKKRTGVSHTTVAH 470
Db 354 -----LLQFNKFEPRVVDLEFKLQKHFDVAKD-----TLN-----VGLRNTTV-- 392
OY 471 PAGNLKSKVTPASTQHDENDTENGDLTNNMKTQVCHVSEISITQRCSSKKTAGLSK 530
Db 393 ---ELSSNETMLKQOYEDIKENLEQKMSSS--KDMAMTINLSV---TQGLINGQVE 444
OY 531 GKTHSASTFYGGESRNGNINIVLSAEOC---QMETENSVLSHAK--VSPAEDHIQIM 586
Db 445 ELLTSSGNITQALVSEKNNTRQELLDASQATKANVASLELVKAYKAEIVQSNEEERIK 504
OY 587 SDLHQSLLPKKKKKQKLEVTREKQTMIDIPMDIVELLAKN-----QHERQLMPTD-- 638
Db 505 HUESERSTLSQKNQIISLGTREAOYEDL---VKLEAKNIEISQISKEBSLTKENEN 561
OY 639 -CSDIRIOSKTTADDDCYIVAAKGSQDVASSVPFDNSQOKLSAQ--STQKELQGHLL 695
Db 562 LSENLKQVQDQLEKLNINLITTK---SNENKISSONELVAKLVSENDLTKORIQOLVEI 618
OY 696 TTQEE---SPHPNQSTQEOQTHLRMEEMVTIAASPLFSHHDOYI-----AEAPTE 745
Db 619 KENEQKDHHTTKLEAFQKNNEQLOKLNVE--VYQLKAHELEDEQNRHLKNCLEKETGEV 676
OY 746 HNGRDAKLTWEQFATTRNSPATCGAQRPRGIQAVDLSTVHWGSSSVYASQPIA 805
Db 677 E-SLSDVTKLQOQVILVLSKQDITAEKLELDQDNLESLEETKNL-----QKRVQ 725
OY 806 PLDRYAEKRVNVOY-----HARNPSTIATMEASKLCD--RNNAGVVLVPKESMPATHL 858
Db 726 SOKRLEQKINLEIKKHKNNKSPKKTQVFTFSPDSPKMAVTTSNLFPNNS-AAHH-- 782
OY 859 RMDPSTIASPPNNGTSSRNOMEQLHNSQYAHNQYKSTSTSYGSNNGKRIPL 912
Db 783 -----SPMKCKPKVDHISKRSINSSKTSKF--NDEPDLSSSSNDDELNTNSPI 830

```

RESULT 13
MAPB_MOUSE
ID MAPB_MOUSE STANDARD: PRT: 2464 AA.
AC P14873;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
GN [Contains: MAP1 light chain LC1].
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DOMAIN.
RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
RX MEDLINE=90094539; PubMed=2480963;
RA Noble M., Lewis S.A., Cowan N.J.;
RT "The microtubule binding domain of microtubule-associated protein
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT and tau."
RL J. Cell Biol. 109:3367-3376(1989).
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC -1- SIMILARITY: TO MAP1A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51396; CAA35761.1; -.
DR PIR: S07549; ORMSPI.
DR MGD: MGI:1306778; Mtap1b.
DR InterPro: IPR000102; MAP1B_neuraxlin.
DR Pfam: PF00414; MAP1B_neuraxlin; 10.
DR PROSITE: PS00230; MAP1B_NEURAXIN; 7.
KM Microtubules; Repeat: Phosphorylation.
FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
FT REPEAT 1874 1890 MAP1B 1.
FT REPEAT 1891 1907 MAP1B 2.
FT REPEAT 1908 1924 MAP1B 3.
FT REPEAT 1925 1941 MAP1B 4.
FT REPEAT 1942 1958 MAP1B 5.
FT REPEAT 1959 1975 MAP1B 6.
FT REPEAT 1993 2009 MAP1B 7.
FT REPEAT 2010 2026 MAP1B 8.
FT REPEAT 2027 2043 MAP1B 9.
FT REPEAT 2044 2060 MAP1B 10.
FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDDBA87 CAC64;

Query Match 3.3%; Score 182.5; DB 1; Length 2464;
Best Local Similarity 18.9%; Pred. No. 0.086;

Matches 175; Conservative 142; Mismatches 357; Indels 251; Gaps 41;
QY 233 PEVQITWHIEVNGAD--QPPSTPKLSEVYLKRNEDENGKEETLVAEQCN-----L 282
DB 528 PPVQVVKLQKORADRESLKPATKPVASKSVKESKETPEVYTKTSOVEKPKVESKEVYL 587
QY 283 TKDNPNSGKERDQVAE--QCNLTKDPPKPVSGQKCEQICNECFEEVYLKRSKSRKTDK 340
DB 588 VKKKPKVTEKSPVTEKEVSSKKEQSPVKA-----EVAEKQAEKSPKPYTKD 635
QY 341 KLMKQOQSHKKRTAADVSDAKLCKRRPKVRLLEIIN-----ANVEDSRDEVHRE 394
DB 636 KVVAKKEIKTKLEEKKEE-----KPKK-----EVKKEDKPYLKKDEKPRKEEYVKE 681
QY 395 NAADPCEDDRTIIPVPEVSMIDIPVSNHTVGEDLKSNNKTRKRYSVVDGSSLMNWL 454
DB 682 IKKKIKKEERREL--KKEVKETP-----LKAQKKEVKK-----EKKKEVKEE 723
QY 455 NGKKKRTGSVHHVTAHPAGNLNKKVPTASTQHDDENDTNGLDJTNHKTQVCOHVEI 514
DB 724 KEPKKEIKKISKDI-----KSTPQSDTK-----KPSALKPKVAKKESTKKEPL 768
QY 515 STQRCSSKKGKTAGLSK-GKTHSAASTRYGGESTRNGONIHVLSADQC-----OMETEN 567
DB 769 AAGLKQKQKGVKVIKKEGKTEAATATVGTAA--TAAVVAAGIAGSPVKELEAER 824
QY 568 SVLSHSAKVSFAE-----HDIQMSDLHQSLPKKKKKKQLEVTRKQTMIDIP 617
DB 825 SLMS-----SPEDLTQDFEELKAEIDVAKDIQD--LELIDEEKLKTQGEAVYVIOKE 878
QY 618 MDIYELLAKNOHERQMLT--ETDC-----SDINRLOSKTADDCVIYAKGVSAYSVF 671
DB 879 TEVSKGSAESDEGITTTEGGECEQPELEPVE-KQGVVD--IEKFDGAGAFEESE 935
QY 672 DTNSQOKSLASQSTOKELQGH--LATLQESPAPHPQNFQSTQEQTHLRMEVNTIAASS 728
DB 936 TGDYEKAETBEAEEPEDEGEDNAGSKSHSPTEDESAKAADVHLKEREVSQGD 995
QY 729 PLESHHDD-----QYIAEAPTEHMGKKAOKLTLWDFKATITN--SPAATGCA- 774
DB 996 RAEDMDVLEKGAEQSEEGEEDKAEDAREGEYEDKTEADRYVVAADKAAEAGVT 1055
QY 775 --QF-----REGIOAVDLSTHVMSSSNVASROPVIALPDRYAERAVNOYHARNF 824
DB 1056 EEOYGYLGTSKKQPGIOS-----PSREP-----ASSHDETLF 1088
QY 825 STIATMENSKLCDRRNAGQVVLVPE-----SMPATHLRMDPSTLAS-- 868
DB 1089 GG-SESEATATADEENREDQ--PEEFATSGYGTSTIEISSEPTPMDEMSTPRDVSDE 1143
QY 869 FPNTGTSRNO-----MESQLNISOYAH-----NOYKGSTF- 900
DB 1144 TNNETESPSQEPVNTIKYESLSLYSOEYSKPAVASFNGLSGSKTDATDGRDYNASASTI 1203
QY 901 -----SYGSNLNGKPLPFEDLSRHQDLHARP-LRPHRQVYL 938
DB 1204 SPSSMEDEKFSKALRDANCSEKELKASAELOIKDVSDBRLSPANSSPISPPSPSTIE 1263
QY 939 GSLQKEIATNMS-----ENCGTQSGYKLGVSIGTISHQNRKREHEALNSGFASKWALQ 994
DB 1264 KTIPLGERSVNFSLPLNEIKVSAEAGEARSVPGV--QAVVEEH-----CASPEEKLE 1314
QY 995 LGSVSSADPLSARNSTIAQSWTRCK 1019
DB 1315 VVSPSQSVTGSAGHTPYQSPPTDEK 1339
RESULT 14
ID P531_HUMAN STANDARD: PRT: 1972 AA.
AC Q12888;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Tumor suppressor p53-binding protein 1 (p53-binding protein 1)
GN TP53BP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE-Skeletal muscle;
RX MEDLINE=98421532; PubMed=9748285;
RA Iwabuchi K., Li B., Massa H.F., Trask B.J., Date T., Fields S.;
RT "Stimulation of p53-mediated transcriptional activation by the
RT p53-binding proteins, 53BP1 and 53BP2."
RL J. Biol. Chem. 273:26061-26068(1998).
RN [2]
RP SEQUENCE OF 946-1972 FROM N.A.
RX MEDLINE=94286584; PubMed=8016121;
RA Iwabuchi K., Bartel P.L., Li B., Marracino R., Fields S.;
RT "Two cellular proteins that bind to wild-type but not mutant p53."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6098-6102(1994).
CC -1- FUNCTION: ENHANCES P53-MEDIATED TRANSCRIPTIONAL ACTIVATION.
CC -1- SUBUNIT: BINDS TO THE CENTRAL DOMAIN OF P53.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. BOTH NUCLEAR AND CYTOPLASMIC
CC IN SOME CELLS.
CC -1- SIMILARITY: CONTAINS 2 BRCT DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF094776; AAC62018.1;
DR EMBL: U09477; AAA21596.1;
DR MIM: 605230;
DR InterPro: IPR001357; BRCT.
DR Pfam: PF00533; BRCT_4.
DR SMART: SM00292; BRCT_2.
DR PROSITE: PS50172; BRCT_2.
DR KMW Nucleic protein; transcription regulation; Activator; Repeat.
FT DOMAIN 1724 1848 BRCT 1.
FT DOMAIN 1864 1964 BRCT 2.
FT DOMAIN 1642 1646 POLY-SER.
FT DOMAIN 1760 1764 POLY-GLU.
SQ SEQUENCE 1972 AA; 213573 MW; 13E2C8A265F9D2A CRC64;

```

Query Match 3.3%; Score 181.5; DI 1; Length 1972;
 Best Local Similarity 18.6%; Pred. NO. 0.073;
 Matches 180; Conservative 135; Mismatches 412; Indels 243; Gaps 40;

```

Db 273 MEAKED---LSAQLMESGLQIKSPEPEVLSTQEDLFDQSN-----KTVSDGCS----- 320
QY 321 EPCEEVLKRRSSKRRKTDKLMKKQOHSKKRPAQADVDAKCRKPKKRVLLSIINA 380
Db 321 -----TPSEBEGCSIASIPATTLHLDLSGOR 348
QY 381 NOVEDSRSDVENHENAAD---PCEDD-RST-----IPVMEVSMIDIPVSNHTVGEDGLK 430
Db 349 SLVQDLSL-----TNSDDLVAPSPDAFRSTPTVPSPPTEGCGROKRPMDTSVLSEGE 403
QY 431 SSKNKTKRRKYSDDVDDGSSLMWMLNGKKRKGSVHTHTVHAPGNLSNKKVYTFATSQHD- 489
Db 404 PRQKTIQSEPEPELENPPLLPSTVSPQASTISOSTPVFPQSL--PIPSQPFSDHI 460
QY 490 -----DENDTENGIDTNMKTVDV---CQHVSEISTOR-----CSSKGTAGLSKGT 533
Db 461 FLPSPLEQSDNGKDKGDMHSSSLTVECSKTSIEPKNSPDLGLSLTGDSCKMLSTS 520
QY 534 HSAASTKYGGESEPT-----RNGONIHVLSADOCOMETENSVLSHSAKVSPAEHDIOMSDL 589
Db 521 EYQSQPKMESLSSHRIDEDGENTQI--ED--TEPSPVL--NSKFPVPAENDSIIM-- 570
QY 590 HEOSLPRKKKKQKLEVTREKQTMIDIPMDIVELLAKNQHROLTETDCSDINRIOSKT 649
Db 571 -----PAQGEVOLSONDDKTKGDDTDTRDISILA-----TGCKG-----RETF 610
QY 650 TADDQCVIYAADGSDYASVPTDINSQOKSLASOSTOKEL-----GCHLALTQESPH- 702
Db 611 VAEDVCIDLTCDSSQAVSPATRSEALSVLDQEMAEIKENHPEEGSSGEVEIEPET 670
QY 703 -----PONFOS-----TOEQHTLRRE-----EMVTIASPLF 731
Db 671 PCSQGEELKEENMSVPLHLSTLTQSGICLOKEMKKESEAMEVETYSIDSP-- 728
QY 732 SHHDQYIAEAPTEWGRKDAKLTWEDQFKATRNSPAATGCAQF--RPIQAVDLTST 788
Db 729 ---OKLALIDDELEH-----KEQEMEE--ATSEDSVVIADVETSPRVVDSCEPLEGV 778
QY 789 HVMGSSSVYASROPYIAD-----LDRYAERAVNOVYHARNPSTIATWASLTCRRANQ 843
Db 779 EKCSQSQSWEDIAPRIECAENRDLTKREKSE--YEDDLKSGTAETD----- 824
QY 844 VLVLPKE---SMPATHLIRMDPSTLASFPNYGTSRNQMSQHLNSQYAHNOYKGSST 900
Db 825 ---PVEDSSQPSLPLVRADPL-----RLDQELQOPQ---TQCKTNSL 863
QY 901 SYGSNL-NGK 909
Db 864 TEDSKMANAK 873

```

RESULT 15
 CENE HUMAN
 ID CENE HUMAN STANDARD; PRT; 2663 AA.
 AC 002224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, last sequence update)
 DT 30-MAY-2000 (Rel. 39, last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENP-E.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schar B.T., Sziliak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis."
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.

```

RX MEDLINE=95196755; PubMed=7889940;
RA Throver D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGSSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARTY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL: Z15005; CAA78727.1; -
DR PIR: S28261; S28261.
DR HSSP: P17119; 3KAR.
DR MIM: 117143; -
DR InterPro: IPR001752; kinesin.
DR Pfam: PF00225; kinesin_1.
DR PRINTS: PRO0380; KINESINHEAVY.
DR SMART: SM00129; KISC: 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN_MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CECF1380C8C8C8B8 CRC64;

```

Query Match 3.3%; Score 181.5; DB 1; Length 2663;
 Best Local Similarity 18.9%; Pred. No. 0.11; Indels 283; gaps 45;
 Matches 185; Conservative 155; Mismatches 357;

```

QY 31 VLELTATPRDAAAEAGVDEPAHQCEHFSIRGVV---ALLOKKDPKFSLSRIFHDOK 86
DB 1822 IOELKANEHQLITLKKDVNTEQKVSEMEQKKQIKDQSLTSLKLEIENLNLAEQELHE-- 1879
QY 87 KDEHKASSSPFVSAAK---FRWDCKSL--DKLTSDNKGTAAPTLPKAKONGTSDGCSI 140
DB 1880 NLEEMK-----SVMKERDNLRAVEETLKLERDQKLESLOETKARDELIQOE----- 1925
QY 141 TFWRSTFVPASVGSQKVPSTQSSQGNADRPSTLPKSVQEGNDKSCNAPSGKNGAEANT 200
DB 1926 -----LKTARMLSKKEKEVVDKIREKISERTIOISDIQKLD-----KSKDELQKKIQ 1973
QY 201 DSPKKDLQGAQNTDVAAVNSDNTSVGALPEVPOITWHIEVN-----GADQPPSTPK 255
DB 1974 ELQKEQLD-----LLRYKED--VNMSHKKINEMEQKKQEPNYLCKCEMDNFOQLTKK 2024
QY 256 LSEY-----VLKRNEDENGTEETLVAAQCNTLPDPPMSGKERD--QVAEQCLUTKDPK 308
DB 2025 LHESLEIRIVAKERDELRIKESLKMEROFIATLREMIARDRONHQVAKPEKRLSDGQ 2084

```

```

QY 309 PVSQKCEQICNEPCEEV--VLKRSKSK-----RKTDKKL----- 342
DB 2085 ---QHMLSELRKESKSRKELLKRTSEBDDHIECLNRLSLDLKEITERRIMKKIKYVLS 2140
QY 343 ---MKQOHSKKRTAQADVSDAKLCRRKPKRVLLSEITIANQVYEDSRSEVHRENAD 398
DB 2141 VYTKIKEEQHNCINKEFMDID-----EVEROKELLIKIQLQDQCDVPSERLR----- 2189
QY 399 PCEDDRSTIIPRMESMDIPVSNRTVGEDGLKSKNKKRKYSDVVDGSS-----LMMWL 454
DB 2190 -----DLKLNQMDLHIE--ILKDFSESEPPSIKTEFOQLSNRKEKETOPLERML 2238
QY 455 -----NGKKRQTSVHNHTVAHPAGNLSNKKVTP--TASPOHDENDENTENDLDTNM 502
DB 2239 NTRPDIKELKNGIOKENDRICO-----VNNFPNNITIMNESTEEFRSATS----- 2287
QY 503 HKTVCQHVSEISTORCSSKGTAGLSKG-----KTHSASTRYGGESTRANGNIHVLSAE 558
DB 2288 -----KEMEDLKLKKEKNEKLFKNYQTLKT-SLASGAQVNPPTQDKKNPHTVSRA 2337
QY 559 DQCM-----ETENSVLSHSAKVPSPAEHDQIMSDL-----HEQSLPKKKK 599
DB 2338 TQLTTEKIRELENSLHEAKESAMHRESKITIKQKLEVTNDIADQAKVHESNCKLEKT 2397
QY 600 KQKLEVTREKQTM-----IDDIPIMDIVEL-LAKNQHROLMETDCSDINRIQSKTT 650
DB 2398 KETIOVLQDKVALGAKPKPEETIEDIKMKLVKIDLEKMKNAKEFEKE----- 2443
QY 651 ADDCVIAAKDGSDYASSVFD--TNSOOKSLASSTOKELQGHIALALTQIESPHPON--- 705
DB 2444 -----ISATKATVEYQKEVIRLRENLRSGQAQDTS-----VISEHNDPOPSNPL 2490
QY 706 -----FOSTOE---QOHLRMEEMVT-----IAASSPLFHHDDQYIAEAPTE 745
DB 2491 TCGGSGGIQONTKALILKSEHIRLEKEISLKOONEOLIKONNELLS--NNQHLSN-EVK 2547
QY 746 HMGKRDAKLTWEQFKATTNSPATCGAOPRPGIAVDLTSTHMGSSSNVYASROPVIA 805
DB 2548 TWKERTLKREAHKO--VTCENSP-----KSPKVTGT-----ASKKKQIT 2584
QY 806 PLDRAEBAVNAQVHARNPSTIATMEASKLC--DRRNAGOVVLYPKESMPATHTLRMPD 863
DB 2585 P-----SQCKERNLQDPV--PRESPKSCFPDSRS-----KSLPSHPVRYFDN 2625
QY 864 STLASFPYGTSSRNQMESQ 883
DB 2626 SSLGICPEVONAGAESYDSQ 2645

```

Search completed: September 16, 2002, 22:35:04
 Job time: 658 sec

THIS PAGE BLANK (uspto)